1-1-1984

Quality control in nonnormal cases a simulation study.

Gon Chen

QUALITY CONTROL IN NONNORMAL CASES

A SIMULATION STUDY

by

Gon Chen

A Thesis

Present in partial fulfillment

of the Requirements for the Degree

Master of Science

in

Industrial Engineering

Lehigh University

1984

QUALITY CONTROL IN NONNORMAL CASES
A SIMULATION STUDY

by

Gon Chen

A Thesis

Present in partial fulfillment

of the Requirements for the Degree

Master of Science

in

Industrial Engineering

Lehigh University

1984

ACKNOWLEDGEMENTS

I dedicate this thesis to my wife.

My special thanks are due to my adviser, Professor J. W. Adams, for his encouragement and patient guidance.

CERTIFICATE OF APPROVAL

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science in Industrial Engineering.

9/28/84

Date

Professor in Charge

chairman
Department of Industrial Engineering

Table of Contents

ABSTRACT

Quality control is worked out for normal distribution. When estimating the location of a Gaussian or normal distribution, the sample mean is well known to be the best estimate according to many criteria. However, the Gaussian model is not appropriate for many situations. For instance, frequently a few large errors infect the data so that the tails of the underlying distribution are heavier than those of Gaussian distribution. In these situations, the sample mean may be no longer a good estimate for the location of symmetry. This thesis will explore some other methods and what happens when the standard theory (assumption of normalty) is used when the data really derive from (1) contaminated normal distribution, (2) slash distribution, (3) cauchy distribution.

Chapter 1

INTRODUCTION

A standard problem in quality control is that of monitoring a process for the purpose of detecting changes in location. The problem investigated in this thesis is that of controlling the location ot a measurable characteristic of an item being mass produced when the assumption of normality is not valid . Location is used instead of mean because for some of the distributions considered, the mean does not exist.

Three questions of interest are the following : First, if control limits are computed, using the normal assumptions , but the plan is implemented in an environment where the distribution of measurements is not normal, what kind of performance is to be expected? Second, how should loction be estimated? And third, could a nonparametric test be used instead of the usual tests used in the normal case?

The answers to the questions are determinted empirically by simulation. This is necessary because a mathematical analysis is impossible.

The determination of the control limits are made from a preliminary sample using the assumption that the data derive from a normal distribution with unknown parameters μ and σ. However, the data actually derive from one of the following distributions:

1. The measurements have a contaminated normal distribution with mean and variance the same as used when simulating the preliminary sample.

2. The measurements have a slash distribution with median equal to the median used when simulating the preliminary sample.

3. The measurements have a cauchy distribution with median equal to the median used when simulating the preliminary sample.

In each case the usual normal tests were used along with the Mann-Whitney and the Huber estimates of location and scale.

In case 1, assuming that the observations of measurable characteric of an item being produced are normally distributed. The standard normal estimate is used to estimate the mean μ and variance σ^2 through the preliminary sample.

$$\mu=(1/N)\sum x_i$$
$$\sigma^2=[1/(n-1)]\sum(x_i-\mu)^2$$

Chapter 2

A DISCUSSION OF THE USUAL QUALITY CONTROL PROBLEM BASED ON THE NORMAL DISTRIBUTION

Suppose that a quality characteristic is measurable on a continuous scale and is known to have a distribution of a specific type , say it is known to have a normal distribution. It is then possible to use a sampling plan based on sample measurments such as the estimated mean of the sample or the estimated mean and standard deviation of the sample. Such plans are called variables sampling plan.

Control chart based on measurments of quality characteristics are often found to be a more economical means of controlling quality than control chart based on attributes. The primary advantage is that the same operating characteristic curve can be obtained with a smaller sample than is required by an attributes plan. The precise measurment required by a variable plan will probably cost more than the simple classification of items required by an attributes plan, but the reduction in sample size may more than offset this extra expense.

If the output of a process derives from a normal distribution, this distribution will be completely described when its mean and standard deviation are known. significant changes in either the mean or the standard deviation are an indication of significant changes in the process, and if specification limits are close to the existing mean , these changes may cause significant changes in the fraction defective.

Suppose that $x_1, x_2, \cdots\cdots, x_n$ are independent random variables and each observations x_i has the normal distribution with parameters mean μ and variance σ^2.

Case 1:

Assuming μ_o and σ^2 are known, for a given sample size n, we construct the upper and lower control limits with which the probability of average of sample falling inside the limits is 0.9974

$UCL=\mu_o+3\sigma/n^{1/2}$ $\qquad$ $LCL=\mu_o-3\sigma/n^{1/2}$

The average of sample $\bar{x}=(1/n)\sum x_i$. Then

$$P_r(\mu_o-3\sigma/n^{1/2}<\bar{x}<\mu_o+3\sigma/n^{1/2})=\int_{\mu_o-3\sigma/n^{1/2}}^{\mu_o+3\sigma/n^{1/2}} \frac{e^{-n/2[(x-\mu)/\sigma]^2}}{(2\pi\sigma^2)^{1/2}}dx$$

$$=\int_{(\mu_o-3\sigma/n^{1/2}-\mu)n^{1/2}/\sigma}^{(\mu_o+3\sigma/n^{1/2}-\mu)n^{1/2}/\sigma} \frac{e^{-z^{1/2}/2}}{(2\pi)^{1/2}}dz$$

If $\mu=\mu_o$, it is mean that there is no shift.

$$P_r(\mu_o-3\sigma/n^{1/2}<\bar{x}<\mu_o+3\sigma/n^{1/2})=\int_{-3}^{3}\frac{e^{-z^{1/2}}}{(2\pi)^{1/2}}dz=0.9974$$

The process is operating satisfactory and it is unlikely that $\bar{x}$ will be outside limits.

If $\mu=\mu_o+\delta$, it is mean that there is a shift of δ in mean. Then

$$P_r(\mu_o-3\sigma/n^{1/2}<\bar{x}<\mu_o+3\sigma/n^{1/2})=\int_{-3-n^{1/2}\delta/\sigma}^{3-n^{1/2}\delta/\sigma}\frac{e^{-z^{1/2}/2}}{(2\pi)^{1/2}}dz$$

The process is operating unsatisfactory and probability of $\bar{x}$ being outside limits is increased.

Case 2:

Assuming the variance is known, but the mean is not known. We have to have a preliminary sample with sample size N to estimate the mean $\bar{X}$, $\bar{X}=(1/N)\sum X_i$

upper control limit $UCL=\bar{X}+k\sigma/n^{1/2}$

lower control limit $LCL=\bar{X}-k\sigma/n^{1/2}$

Where n is sample size tor monitoring the process. Supposing that $\bar{x}_i$ is the i^{th} sample average value. Then

$$\bar{x}_i=(1/n)\sum x_j$$

The probability of $\bar{x}_i$ fall whithin limits is:

$$P_r(\bar{X}-k\sigma/n^{1/2}<\bar{x}_i<\bar{X}+k\sigma/n^{1/2}) \quad \text{or}$$

$$P_r(-k\sigma/n^{1/2}<\bar{x}_i-\bar{X}<k\sigma/n^{1/2})$$

$\bar{x}_i-\bar{X}$ has normal distribution with parameters $\mu-\mu_o$ and $\sigma^2*(1/n+1/N)$. So

$$P_r(\bar{x}_i \text{ within limits})$$

$$=\int_{-k\sigma/n^{1/2}}^{k\sigma/n^{1/2}} \frac{e^{-\frac{1}{2}\left(\frac{\bar{x}_i-\mu+\mu_o}{[(1/n+1/N)\sigma^2]^{1/2}}\right)^2}}{[2\pi(1/n+1/N)\sigma^2]^{1/2}}\,dx$$

$$=\int_{(-k\sigma/n^{1/2}-\mu+\mu_o)[nN/(n+N)]^{1/2}1/\sigma}^{(k\sigma/n^{1/2}-\mu+\mu_o)[nN/(n+N)]^{1/2}1/\sigma} \frac{e^{-z^2/2}}{(2\pi)^{1/2}}dz$$

$$=\int_{-k[N/(n+N)]^{1/2}-(\mu-\mu_o)[nN/(n+N)]^{1/2}1/\sigma}^{k[N/(n+N)]^{1/2}-(\mu-\mu_o)[nN/(n+N)]^{1/2}1/\sigma} \frac{e^{-z^2/2}}{(2\pi)^{1/2}}dz$$

When $\mu=\mu_o$, the probability that $\bar{x}_i$ inside the limits is

$$\int_{-k[N/(n+N)]^{1/2}}^{k[N/(n+N)]^{1/2}} \frac{e^{-z^2/2}}{(2\pi)^{1/2}}dz=0.9974$$

Then $\quad k[N/(n+N)]^{1/2}=3 \quad$ k can be calculated.

When $\mu=\mu_o+\delta$, the probability that $\bar{x}_i$ inside the limits is

$$\int_{-k[N/(n+N)]^{1/2}-\delta/\sigma[nN/(n+N)]^{1/2}}^{k[N/(n+N)]^{1/2}-\delta/\sigma[nN/(n+N)]^{1/2}} \frac{e^{-z^2/2}}{(2\pi)^{1/2}}dz$$

$$=\int_{-3-\delta/\sigma[nN/(n+N)]^{1/2}}^{3-\delta/\sigma[nN/(n+N)]^{1/2}} \frac{e^{-z^2/2}}{(2\pi)^{1/2}}dz$$

When unknown mean and known variance, we have to estimate the mean and calculate the k value , then, get the probability of of sample within limit. When N value become larger the probablity will be more accurate.

Case 3:

Assuming that we know mean, but the variance is unknown. We have to have a preliminary sample with sample size N to estimate the variance.

$s_o{}^2=1/N\sum(X_i-\mu_o)^2$ $\quad$ $s_o{}^2$ is the estimate of variance

$\bar{x}_i=1/n\sum x_{ij}$

$s_i=1/(n-1)\sum(x_{ij}-\bar{x}_i)^2$ $\quad$ variance for i^{th} sample with size n

$$s^2=\frac{Ns_o{}^2+(n-1)s_i{}^2}{n+N-1}$$

where

$$\frac{(n+N-1)s^2}{\sigma^2}$$ has chi-square dist. with n+N-1 d.f.

$$t=\frac{n^{1/2}(\bar{x}-\mu_o)}{s}$$ has t-dist. with n+N-1 d.f. when $\mu=\mu_o$

Let $H_o:\mu=\mu_o$ and $H_a:\ \mu\neq\mu_o$

When $\mu=\mu_o$, H_o is true , then

$P_r(-t_{\alpha/2}<t<t_{\alpha/2})=1-\alpha$

The probability of reject H_o when it is true is α.

When $\mu=\mu_o+\delta$ $\quad\delta\neq0$

$n^{1/2}(\bar{x}-\mu_o)/s$ has noncentral t-distribution with n+N-1 d.f. and noncentral parameter $n^{1/2}\delta/\sigma$.

The noncentral t can be evaluated as a function of $n^{1/2}\delta/\sigma$.

$P_r(-t_{\alpha/2}<n^{1/2}(\bar{x}-\mu_o)/s<t_{\alpha/2}\quad|\mu=\mu_o+\delta)$

Now, if there is a shift in μ, say δ, the probability of accepting can be found through noncentral t-distribution table.

Chapter 3

THE ALTERNATIVE ASSUMPTIONS OF DISTRIBUTION TO BE INVESTIGATED

3.1 CONTAMINATED NORMAL DISTRIBUTION

The contaminated normal distribution represents a mixture of observations taken from a standard Gaussian with probability 1-p and from the wilder Gaussian (with scale H and the same center) with probability p. The distribution function for such a contaminated normal distribution is

$$F(x)=(1-p)G(x)+pG(x/H)$$

Where G(x) is the cumulative distribution function for the standard Gaussian. The density function of the contaminated normal distribution is

$$f(x)=(1-p)\frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{(2\pi\sigma^2)^{1/2}} + p\frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{H\sigma}\right)^2}}{(2\pi H^2\sigma^2)^{1/2}}$$

If j values are contaminated, the average value $\bar{x}$ of a sample with sample size n has the normal distribution with parameter μ and $(jH^2+n-j)\sigma^2$. Where μ is the mean value.

The density function is

$$f(x)=\sum_{j}\binom{n}{j}p^{j}(1-p)^{n-j}\frac{e^{-\frac{n}{2}\left(\frac{x-\mu}{[(jH^{2}+n-j)\sigma^{2}]^{1/2}}\right)^{2}}}{[2\pi(jH^{2}+n-j)\sigma^{2}]^{1/2}}$$

The cumulative density function is

$$P_r(X<x)=\sum_{j}\binom{n}{j}p^{j}(1-p)^{n-j}\int_{-\infty}^{n^{1/2}(x-\mu)/[(jH^{2}+n-j)\sigma^{2}]^{1/2}}\frac{e^{-z^{2}/2}}{(2\pi)^{1/2}}dz$$

We can generate contaminated distribution values by simulation using the given parameter values $n,\mu,\sigma,p,$ and H.

3.2 SLASH DISTRIBUTION

The slash distribution is defined by a standard Gaussian random variable divided by an independent uniform random variable on the interval (0,1).

When X has the normal distribution with parameter μ and σ^2, Y has the uniform distribution on (0,1), and $Z=X/Y$, X and Y are independent. Then

$$P_r(Z\le t)=P_r(X/Y\le t)=P_r(X\le tY)$$

$$=\int_0^1 P_r(X\le tY \mid Y=y)\,dP_r(Y\le y)$$

$$=\int_0^1 P_r(X\le ty \mid Y=y)\,f_Y(y)\,dy$$

$$=\int_0^1 P_r(X\le ty)\,dy$$

$$=\int_0^1 G(ty)\,dy$$

$$=\int_0^1\int_{-\infty}^{ty} \frac{e^{-\mu^2/2}}{(2\pi)^{1/2}}\,d\mu\,dy$$

Let $s=\dfrac{u}{y}$ $ds=\dfrac{1}{y}\,du$

So

$$P_r(Z\le t)=\int_0^1 \int_{-\infty}^{t} y\frac{e^{-(1/2)s^2y^2}}{(2\pi)^{1/2}}\,ds\,dy$$

The density function of slash distribution should be

$$f(t)=\int_0^1 y\frac{e^{-(1/2)t^2y^2}}{(2\pi)^{1/2}}\,dy$$

$$E\left(\frac{X}{Y}\right)=E(X)*E\left(\frac{1}{Y}\right)=u\int_0^1 \frac{1}{y}\,dy=\infty$$

$$E\left(\frac{X}{Y}\right) \text{ and } var\left(\frac{X}{Y}\right) \text{ do not exist.}$$

3.3 CAUCHY DISTRIBUTION

The cauchy distribution is an example of a distribution for which no moments exist. The probability density function is

$$f(x) = \frac{1}{\pi B} * \frac{1}{1+[(x-T)/B]^2}$$

Where $-\infty<x<\infty$, $B>0$, and $-\infty<T<\infty$. The cumulative probability distribution is

$$P_r(X<x) = \frac{1}{\pi} \arctan\left(\frac{x-T}{B}\right) + \frac{1}{2}$$

The median and mode of this distribution are at $x=T$. For $B=1$, and $T=0$, cauchy distribution is also a robust t-distribution with 1 degree of freedon.

Contaminated normal, slash, cauchy, can be represented on the Normal/ Independent form of the sampling distributions.

An observation x can be generated on the computer as Z/Y, where Z is unit Normal and Y is generated independently of Z. Table 1 shows that different independent divisors generate different distribution.

Table 1, The exhibition of form Normal/Independent

Sampling Distribution $X=Z/Y$	Independent Distribution
Normal	Degenerate $Y=1$
Contaminated normal $F(X)=pG(\frac{X}{H})+(1-p)G(X)$	$Y=\begin{cases} H & \text{with prob. } p \\ 1 & \text{with prob. } 1-p \end{cases}$
Slash	$f(Y)=1/K$, $K>0$, $0<Y<1$
Cauchy	Half-normal $F(Y)=\begin{cases} 2[G(Y)-0.5], & Y>0 \\ 0 , & Y\leq 0 \end{cases}$

Chapter 4

CASES FOR ANALYZING BY SIMULATION

4.1 STANDARD PROCESS

If we assumed that the output of a process comes from a normal distribution, the mean and standard deviation are considered to be important measures of the distribution.

Significant changes in either the mean or the standard deviation are an indication of significant changes in the process,and if specification limits are close to the existing mean, these changes may cause significant changes in the fraction defective.

When a preliminary sample with size N was taken, we can estimate the parameter values μ' and σ'^2 with formula

$$\mu'=(1/N)\sum x_i$$

$$\sigma'^2=[1/(N-1)]\sum(x_i-\mu')^2$$

Assuming the significant level $\alpha=0.01$. The formulas for the control limits are.

$$UCL=\mu'+2.57\sigma'/n^{1/2}.$$

$$LCL=\mu'-2.57\sigma'/n^{1/2}.$$

A sample of n items is taken from the process every so often and a quality measurement made of each item. The average of these measurements is then computed. When a point falls outside the control limits, the process is deemed to be out of control with respect to its central tendency.

4.2 NONPARAMETRIC METHOD - MANN AND WHITNEY U TEST

The Mann - Whitney U Test is based on the idea that the particular pattern exhibted when m X random variables and n Y random variables are arranged together in increasing order of magnitude provides information about the relationship between their populations. The Mann - Whitney criterion is based on the magnitudes of the Y's in relation to the X's, that is, the positions of the Y's, in the combined ordered sequence. A sample pattern of arrangement where most of the Y's are greater than most of the X's, or vice versa, or both, would be evidence against a radom mixing and thus tend to discredit the null hypothesis of identical distribution.

We could count the number of Y observations that precede each X observations and use the sum of the counts, U_Y, as the U statisic, or vice versa. In either case, very large or small values of U will imply a separation of the ordered X and Y observations, and will provide evidence to indicate a difference(a shift in location) between the population distribution for X and Y.

U_X=nm+[m(m+1)/2]−W_X

U_Y=nm+[n(n+1)/2]−W_Y

Where $U_X+U_Y=nm$ and W_X and W_Y are the rank sums for samples X and Y, respectively.

When m and n are large enough ($m>10, n>10$), the asymptotic probability distribution can be used.

Since U is the sum of identically distributed (through dependent) random variable, a generalization of the central limit theorem allows us to conclude that the null distribution of the standardized U approaches the standard normal. To make use of this approximation the expected mean and variance of U under the null hypothesis must be determined.

$E(U)=nm/2 \qquad Var(U)=nm(n+m+1)/12$

$Z=[U-E(U)]/[nm(n+m+1)/12]^{1/2}$

When we have a shift in mean, $\mu=\mu_0+\delta$, assuming the significant level $\alpha=0.01$, then, We can find the point out side the limits if $Z>2.57$, or $Z<-2.57$.

4.3 USE M- ESTIMATOR TO DETECT THE SHIFT IN LOCATION

Certainly the method of least squares and generalization of it have served us well for many years. However, it is recognized that 'outliers', which arise from heavy tailed distributions or are simply bad data points due to errors, have an unusually large influence on the least squares estimators. The robust methods have been created to modify least squares schemes so that the outliers have much less influence on the final estimates. One of the most satisfying robust procedures is that given by a modification of the principle of maximum likelihood. It should be mentioned that Huber provides an excellent summary ot many of the mathmetical aspects of robustness. So we choose Huber's function.

Let $x_1, x_2, \cdots\cdots, x_n$ be a random sample that arises from a distribution with density $f(x-t)$ of the coutinuous type, where t is a location parameter.

The logarithm of the likelihood function is

$$\ln L(t)=\sum \ln f(x_i-t)=-\sum r(x_i-t)$$

Where $r(x)=-\ln f(x)$. In maximum likelihood we wish to maximize $\ln L(t)$, or minimize $\sum r(x_i-t)$. Suppose that this minimization can be achieved by differentiating, that is finding the appropriate t that satisfies

$$\sum \Psi(x_i - t) = 0$$

Where $\Psi(x)=r'(x)=-f'(x)/f(x)$. The solution of this equation is called the maximum likelihood or M-estimator of t and is denoted by t.

Using a more technical definition of robustness, Huber derived the following robust r and Ψ functions.

$$r(x)=\begin{cases} x^2/2 , & |x|\le a \\ a|x|-a^2/2 , & |x|>a \end{cases}$$

$$\Psi(x)=\begin{cases} -a , & x<-a \\ x , & |x|\le a \\ a , & x>a \end{cases}$$

To create a scale invariant version of the M-estimator. We find the solution t of

$$\sum \Psi[(x_i-t)/s]=0$$

Where $s=cs_n$, the robust estimator of scale.

$s_n=\text{median}|x_i-\text{median } x_i|$,MAD,

the auxiliary estimator of scale.

$c=1/0.6745$, the tuning constant, the median absolute deviation may be normalized by dividing the value 0.6745 for the standard Gaussian distribution.

We use t_o=sample median as the first guess and by Newton's method

$$t_j=t_{j-1}+\frac{s\sum Y[(x_i-t_{j-1})/s]}{\sum Y'[(x_i-t_{j-1})/s]}$$

Where $\sum Y'[(x_i-t_{j-1})/s]$ counts the number of items that enjoy $|x_i-t_{j-1}|/s< a$

The Y-function ot a Huber estimator, shown in FIGURE 1, is linear in the center and constant in the tail.

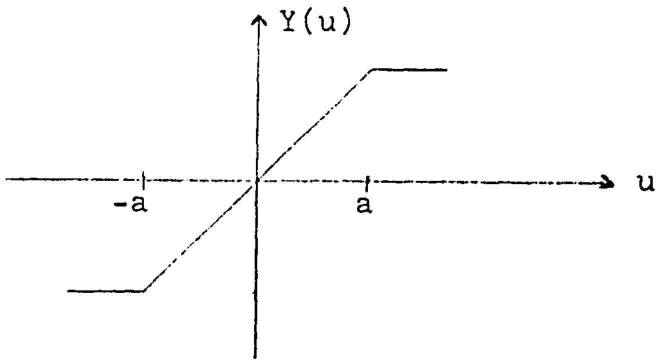

Figure 1. Y-function of Huber's estimator of location

The mean and median are extreme cases in Huber's family. If the sample actually comes from normal distribution, most of the item would enjoy the property that $|x_i-t|/s<1.5$. And if this value more large ,then $\Psi[(x_i-t)/s]=(x_i-t)/s$, and $\sum\Psi[(x_i-t)/s]=0$ has the solution $t=\bar{x}$, as desired in normal case.

In practice for quality control, when we find the location and scale estimates. We can construct the control limits. Assuming significant level $\alpha=0.01$ then

$UCL=t+2.57s/n^{1/2}$

$LCL=t-2.57s/n^{1/2}$

A sample of n items is taken from the process every so often and a quality measurement made of each item. The Huber's estimate of location t is then computed. So long as a point falls outside the control limits , the process is deemed to be a shift in location.

In Huber's family , there is different tuning constant 'a' on different population distribution. TABLE 2 and TABLE 3 show that the tuning constant 'a' should be 1.0 or 1.3 on contaminated distribution (when p=.05, H=2 and 3 respectively). TABLE 3 AND TABLE 4 show that the tuning constant 'a' should be 2.0 on cauchy distribution and slash distribution. The criterion is based on the power (No. of out of control in 100 samples) of tuning constant 'a' from [0.7, 1.0, 1.3, 1.5, 2.0] on a specified population distribution. Those data are collected by computer simulation.

Chapter 5

SIMULATION MODEL AND EXPERIMENTAL DESIGN

In the development of a simulation model to monitor a process for the purpose of detecting changes in the location of distribution. At startup a preliminary sample with sample size N is needed. In this thesis, we use N=600, and I.M.S.L. subroutine to work on Cyber 730 to generate data by simulation.

5.1 THE THREE DIFFERENT CASES

5.1.1 Simulation under standard process

Under this case we consider that observations come from a normal distribution. From the preliminary 600 values, we estimate the parameter values $\mu'=(1/600)\sum x_i$, and $\sigma'^2=[1/(600-1)]\sum(x_i-\mu')^2$. Then the control limit can be computed

$$UCL=\mu'+2.57\sigma'/n^{1/2}$$

$$LCL=\mu'-2.57\sigma'/n^{1/2}$$

The samples for monitoring the process with sample size n=25, can be generated and the sample value $\bar{x}$, the average of the sample, can be computed. If $\bar{x}>UCL$, or, $\bar{x}<LCL$ the point be considered out of control. If a shift in location happens, the number of points out of control in large number of samples should be considered the power of this case.

5.1.2 Simulation under Mann-Whitney U Test

Under this case, the preliminary 600 observations to be considered as X's, and the samples for monitoring the process with sample size n=25 as Y's are arranged together in incereasing order of magnitude. When a difference (a shift in location) between the population distribution for X and Y happens, the information will be provided by

$$U_Y=600*25+25*(25+1)/2-W_Y$$

Where W_Y is the rank sum of sample Y .

The expected mean $E(U)=600*25/2=7500$

The variance $Var(U)=600*25*(600+25+1)/12=782500$

$$Z=(U_Y-7500)/(782500)^{1/2}$$

If $\alpha=0.01$ and there is a shift happens, then $Z>2.57$ or $Z<-2.57$ means that a point outside the limits. The number of point out of limit in large number of samples should be considered the power of this case.

5.1.3 Simulation under Huber estimate of location and scale

Under this case, We find the Huber estimate of location t and estimate of scale s by the following equations.

Let

M=median $(x_1, x_2, \cdots\cdots, x_n)$

s=median $(|x_1-M|, |x_2-M|, \cdots\cdots, |x_n-M|)/0.6745$

If $\hat{t}$ is the estimate of location. It satisfyes the equation

$$\sum Y[(x_i-\hat{t})/s]=0$$

Where

$$Y[(x_i-\hat{t})/s]=\begin{cases} -a , & (x_i-\hat{t})/s<-a \\ (x_i-\hat{t})/s , & |x_i-\hat{t}|/s\leq a \\ a , & (x_i-\hat{t})/s>a \end{cases}$$

Use Newton's method to solve the equation as follow:

$$\hat{t}_j=\hat{t}_{j-1}+s\frac{\sum Y[(x_i-\hat{t}_{j-1})/s]}{\sum Y'[(x_i-\hat{t}_{j-1})/s]} \qquad ,\hat{t}_o=M$$

Where $\sum Y'[(x_i-\hat{t}_{j-1})/s]$ is the number of items that enjoy $|x_i-\hat{t}_{j-1}|/s<a$.

From the preliminary 600 values, we can estimate the location T and scale S of the process assuming this proess under control, then the control limits can be computed by equation:

$$UCL=T+2.57S/n^{1/2}$$

$$LCL=T-2.57S/n^{1/2}$$

The sample for monitoring the process with sample size n=25 can be generated and the sample value t, the Huber estimate of location of this sample, can be computed. If t>UCL, or, t<LCL, the point is considered to be out of limit and the number of points out of limit in a large number of sample is considered to be the power of this case.

The tuning constant 'a' should be different with respect to different population distibution. TABLES from 2 to 5, show the smulation result of finding appropriate constant 'a' for different population distribution. When contaminated distribution, the value 'a' should be 1.0 or 1.3 for H=3 or 2 respectively. When slash or cauchy distribution, the value 'a' should be 2.0.

Chapter 6

RESULTS OF SIMULATION

Table 2 , Power of rejection,each seed has 100 samples with sample size 25. Population distribution is contaminated distribution with p=.05, H=2 The shift δ=.5

seed	No. of rejection by Huber				
	constant value 'a'				
	.7	1.0	1.3	1.5	2.0
7118863	48	46	45	44	41
6512521	69	68	68	66	62
2412827	46	48	48	51	49
3341985	34	33	34	35	36
4587736	48	50	49	51	50
5714837	39	38	40	39	40
7682513	38	39	39	34	34
2914508	61	67	71	70	68
1983451	36	38	35	34	35
Total	419	427	429	424	415

Table 3 , Power of rejection, each seed has 100 samples with sample size 25. Population distribution is contaminated distribution with p=.05, H=3, The shift δ=.5

seed	No. of rejection by Huber				
	constant value 'a'				
	.7	1.0	1.3	1.5	2.0
7118863	47	44	43	41	40
6512521	67	67	64	63	59
2412827	46	46	48	49	49
3341985	35	33	31	32	36
4587736	48	49	48	50	49
5714837	39	38	39	39	40
7682513	40	39	33	33	33
2914508	60	67	70	69	67
1983451	36	37	34	32	33
Total	418	420	410	408	406

Table 4 , Power of rejction, each seed has 100 samples with sample size 25. Population distribution is cauchy distribution. The shift δ=.5

Seed	Power(No. of rejection) Constant value 'a'		
	1.3	1.5	2.0
7118863	17	18	21
6512521	40	39	40
2412827	34	37	37
3341985	25	25	26
4587736	29	31	38
5714837	15	16	16
7682513	14	17	19
2914508	14	16	20
1983451	26	27	32
8686368	31	31	32
Total	245	257	281

Table 5 , Power of rejection, each seed has 100 samples with sample size 25. Population distribution is slash distribution. The shift δ=.5

Seed	Power(No. of rejection)		
	Constant value 'a'		
	1.3	1.5	2.0
7118863	11	13	16
6512521	17	16	17
2412827	22	23	25
3341985	5	6	9
4587736	17	18	16
5714837	9	9	11
7682513	18	19	18
2914508	37	38	38
1983451	4	5	9
8686368	6	6	7
Total	146	153	166

Table 6, Comparision of contaminated distribution with H=3 ,p=.05

seed=7118863, sample size n=5

Location shift δ	A	B	C	A-B	A-C	B-C
.00	4.0	1.0	5.0	3	-1	-4
.10	2.0	.0	3.0	2	-1	-3
.20	2.0	2.0	1.0	0	1	1
.30	1.0	.0	2.0	1	-1	-2
.40	2.0	1.0	7.0	1	-5	-6
.50	10.0	7.0	7.0	3	3	0
.60	15.0	13.0	19.0	2	-4	-6
.70	11.0	16.0	15.0	-4	-4	1
.80	14.0	11.0	17.0	3	-3	-6
.90	14.0	14.0	24.0	0	-10	-10
1.00	34.0	32.0	48.0	2	-14	-18
1.10	33.0	28.0	43.0	5	-10	-15
1.20	49.0	44.0	54.0	5	-5	-10
1.30	57.0	50.0	65.0	7	-8	-15
1.40	59.0	50.0	69.0	9	-10	-19
1.50	68.0	62.0	73.0	6	-5	-11

A:Power of $\bar{x}$(No. of rejections in 100 samples)
B:Power of Mann-Whitney
C:Power of Huber
m_A=42.875 , m_B=44.438 , m_C=47.063 , the average of A,B,C respectively.

Table 7, Comparision of contaminated distribution with H=3, p=.05

seed=7118863, sample size n=10

Location shift δ	A	B	C	A-B	A-C	B-C
.00	3.0	2.0	3.0	1	0	-1
.10	2.0	3.0	3.0	-1	-1	0
.20	1.0	.0	5.0	1	-4	-5
.30	7.0	8.0	9.0	-1	-2	-1
.40	7.0	12.0	9.0	-5	-2	3
.50	22.0	20.0	26.0	2	-4	-6
.60	26.0	26.0	27.0	0	-1	-1
.70	28.0	33.0	33.0	-5	-5	0
.80	42.0	39.0	46.0	3	-4	-7
.90	56.0	60.0	66.0	-4	-10	-6
1.00	59.0	65.0	69.0	-6	-10	-4
1.10	68.0	77.0	78.0	-9	-10	-1
1.20	89.0	87.0	93.0	2	-4	-6
1.30	86.0	87.0	93.0	-1	-7	-6
1.40	93.0	94.0	94.0	-1	-1	0
1.50	97.0	98.0	99.0	-1	-2	-1

A:Power of $\bar{x}$(No. of rejections in 100 samples)
B:Power of Mann-Whitney
C:Power of Huber
m_A=42.875 , m_B=44.438 , m_C=47.063 , the average of A,B,C respectively.

Table 8, Comparision of contaminated distribution with H=3, p=.05

seed=7118863, sample size n=15

Location shift δ	A	B	C	A-B	A-C	B-C
.00	2.0	2.0	3.0	0	-1	-1
.10	2.0	1.0	3.0	1	-1	-2
.20	8.0	9.0	8.0	-1	0	1
.30	5.0	8.0	7.0	-3	-2	1
.40	18.0	21.0	23.0	-3	-5	-2
.50	17.0	22.0	23.0	-5	-6	-1
.60	45.0	48.0	51.0	-3	-6	-3
.70	50.0	63.0	59.0	-13	-9	4
.80	57.0	69.0	70.0	-12	-13	-1
.90	70.0	79.0	79.0	-9	-9	0
1.00	81.0	86.0	87.0	-5	-6	-1
1.10	83.0	88.0	89.0	-5	-6	-1
1.20	95.0	98.0	98.0	-3	-3	0
1.30	98.0	98.0	99.0	0	-1	-1
1.40	99.0	99.0	100.0	0	-1	-1
1.50	99.0	100.0	100.0	-1	-1	0

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_A=51.83 , m_B=55.69 , m_C=56.19 , the average of A,B,C respectively.

Table 9, Comparision of contaminated distribution with H=3, p=.05

seed=7118863, sample size n=20

Location shift δ	A	B	C	A-B	A-C	B-C
.00	1.0	1.0	1.0	0	0	0
.10	6.0	7.0	6.0	-1	0	1
.20	10.0	11.0	11.0	-1	-1	0
.30	12.0	16.0	18.0	-4	-6	-2
.40	21.0	31.0	28.0	-10	-7	3
.50	32.0	38.0	38.0	-6	-6	0
.60	49.0	55.0	53.0	-6	-4	2
.70	59.0	68.0	71.0	-9	-12	-3
.80	73.0	81.0	82.0	-8	-9	-1
.90	83.0	90.0	91.0	-7	-8	-1
1.00	94.0	98.0	99.0	-4	-5	-1
1.10	95.0	98.0	98.0	-3	-3	0
1.20	100.0	100.0	100.0	0	0	0
1.30	99.0	100.0	100.0	-1	-1	0
1.40	100.0	100.0	100.0	0	0	0
1.50	100.0	100.0	100.0	0	0	0

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_A=58.375 , m_B=62.125 , m_C=62.25 , the average of A,B,C respectively.

Table 10, Comparision of contaminated distribution with H=3, p=.005

seed=7118863, sample size n=25

Location shift δ	A	B	C	A−B	A−C	B−C
.00	4.0	2.0	4.0	2	0	−2
.10	.0	2.0	3.0	−2	−3	−1
.20	5.0	12.0	11.0	−7	−6	1
.30	13.0	25.0	16.0	−12	−3	9
.40	30.0	41.0	31.0	−11	−1	10
.50	37.0	48.0	41.0	−11	−4	7
.60	56.0	71.0	66.0	−15	−10	5
.70	76.0	88.0	86.0	−12	−10	2
.80	88.0	91.0	91.0	−3	−3	0
.90	90.0	96.0	96.0	−6	−6	0
1.00	97.0	99.0	99.0	−2	−2	0
1.10	98.0	100.0	100.0	−2	−2	0
1.20	99.0	100.0	100.0	−1	−1	0
1.30	100.0	100.0	100.0	0	0	0
1.40	100.0	100.0	100.0	0	0	0
1.50	100.0	100.0	100.0	0	0	0

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_A=62.06 , m_B=67.19 , m_C=65.25 , the average of A,B,C respectively.

Taole 11, Comparision of slash distribution with seed=7118863

sample size n=5

LOcation shift δ	A	B	C	B-C
.00	5.0	1.0	10.0	-9
.10	2.0	.0	11.0	-11
.20	5.0	1.0	14.0	-13
.30	4.0	.0	10.0	-10
.40	3.0	1.0	11.0	-10
.50	4.0	3.0	13.0	-10
.60	5.0	5.0	22.0	-17
.70	3.0	1.0	10.0	-9
.80	3.0	2.0	11.0	-9
.90	6.0	4.0	15.0	-11
1.00	4.0	8.0	23.0	-15
1.10	7.0	3.0	13.0	-10
1.20	4.0	8.0	25.0	-17
1.30	5.0	10.0	17.0	-7
1.40	2.0	5.0	18.0	-13
1.50	5.0	6.0	16.0	-10
1.60	4.0	11.0	32.0	-21
1.70	4.0	14.0	34.0	-20
1.80	3.0	20.0	33.0	-13
1.90	3.0	23.0	35.0	-12
2.00	3.0	27.0	46.0	-19
2.10	2.0	21.0	37.0	-16
2.20	4.0	30.0	37.0	-7
2.30	2.0	31.0	48.0	-17
2.40	1.0	34.0	57.0	-23
2.50	2.0	36.0	56.0	-20
2.60	3.0	25.0	57.0	-32

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=12.2 , m_C=26.3 , the average of B and C respectively.

Table 12, Comparision of slash distribution with seed=7118863

sample size n=10

Location shift δ	A	B	C	B-C
.00	11.0	2.0	15.0	-17
.10	7.0	.0	8.0	-8
.20	4.0	4.0	12.0	-8
.30	11.0	4.0	11.0	-7
.40	9.0	8.0	18.0	-10
.50	2.0	4.0	12.0	-8
.60	5.0	15.0	25.0	-10
.70	5.0	15.0	24.0	-9
.80	11.0	9.0	21.0	-12
.90	5.0	16.0	32.0	-16
1.00	5.0	21.0	42.0	-21
1.10	11.0	28.0	37.0	-9
1.20	5.0	29.0	41.0	-12
1.30	11.0	33.0	40.0	-7
1.40	5.0	40.0	52.0	-12
1.50	10.0	45.0	63.0	-18
1.60	6.0	51.0	67.0	-16
1.70	11.0	44.0	66.0	-22
1.80	5.0	61.0	78.0	-17
1.90	8.0	54.0	66.0	-12
2.00	9.0	68.0	79.0	-11
2.10	7.0	69.0	78.0	-9
2.20	7.0	78.0	93.0	-15
2.30	7.0	77.0	86.0	-9
2.40	11.0	81.0	94.0	-13
2.50	6.0	85.0	90.0	-5
2.60	7.0	89.0	96.0	-7

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=38.15 , m_C=49.85 , the average of B and C respectively.

Table 13, Comparision of slash distribution with seed=7118863

sample size n=15

Location shift δ	A	B	C	B-C
.00	11.0	2.0	15.0	-13
.10	7.0	.0	8.0	-8
.20	4.0	4.0	12.0	-8
.30	11.0	4.0	11.0	-7
.40	9.0	8.0	18.0	-10
.50	2.0	4.0	12.0	-8
.60	5.0	15.0	25.0	-10
.70	5.0	15.0	24.0	-9
.80	11.0	9.0	21.0	-13
.90	5.0	16.0	32.0	-16
1.00	5.0	21.0	42.0	-21
1.10	11.0	28.0	37.0	-9
1.20	5.0	29.0	41.0	-12
1.30	11.0	33.0	40.0	-7
1.40	5.0	40.0	52.0	-12
1.50	10.0	45.0	63.0	-18
1.60	6.0	51.0	67.0	-16
1.70	11.0	44.0	66.0	-22
1.80	5.0	61.0	78.0	-17
1.90	8.0	54.0	66.0	-12
2.00	9.0	68.0	79.0	-11
2.10	7.0	69.0	78.0	-9
2.20	7.0	78.0	93.0	-15
2.30	7.0	77.0	86.0	-9
2.40	11.0	81.0	94.0	-13
2.50	6.0	85.0	90.0	-5
2.60	7.0	89.0	96.0	-7

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=36.2 , m_C=49.85 , the average of B and C respectively.

Table 14, Comparision of slash distribution with seed=7118863

sample size n=20

Location shift δ	A	B	C	B-C
.00	8.0	1.0	6.0	-5
.10	5.0	5.0	14.0	-9
.20	7.0	6.0	15.0	-9
.30	10.0	6.0	16.0	-10
.40	3.0	4.0	14.0	-10
.50	4.0	12.0	23.0	-11
.60	9.0	11.0	22.0	-11
.70	7.0	12.0	20.0	-8
.80	12.0	22.0	34.0	-12
.90	9.0	31.0	35.0	-4
1.00	9.0	36.0	44.0	-8
1.10	5.0	33.0	40.0	-7
1.20	2.0	39.0	51.0	-12
1.30	11.0	42.0	50.0	-8
1.40	12.0	48.0	59.0	-11
1.50	14.0	67.0	76.0	-9
1.60	8.0	62.0	72.0	-10
1.70	6.0	68.0	80.0	-12
1.80	8.0	75.0	80.0	-5
1.90	6.0	76.0	85.0	-9
2.00	8.0	84.0	88.0	-4
2.10	4.0	83.0	88.0	-5
2.20	4.0	87.0	92.0	-5
2.30	8.0	90.0	92.0	-2
2.40	12.0	94.0	98.0	-4
2.50	16.0	93.0	96.0	-3
2.60	8.0	98.0	99.0	-1

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=47.592 , m_C=53.148 , the average of B and C respectively.

Table 15, Comparision of slash distribution with seed=7118863

sample size n=25

Location shift δ	A	B	C	B-C
.00	5.0	2.0	8.0	-6
.10	9.0	2.0	15.0	-13
.20	7.0	7.0	14.0	-7
.30	6.0	5.0	13.0	-8
.40	3.0	7.0	17.0	-10
.50	11.0	14.0	19.0	-5
.60	9.0	12.0	19.0	-7
.70	7.0	23.0	29.0	-6
.80	5.0	26.0	35.0	-9
.90	9.0	34.0	40.0	-6
1.00	8.0	36.0	42.0	-6
1.10	9.0	41.0	48.0	-7
1.20	10.0	59.0	64.0	-5
1.30	8.0	59.0	63.0	-4
1.40	5.0	73.0	73.0	0
1.50	11.0	76.0	78.0	-2
1.60	10.0	87.0	91.0	-4
1.70	10.0	84.0	84.0	0
1.80	6.0	91.0	94.0	-3
1.90	8.0	96.0	97.0	-1
2.00	7.0	90.0	91.0	-1
2.10	12.0	95.0	96.0	-1
2.20	10.0	96.0	97.0	-1
2.30	5.0	97.0	99.0	-2
2.40	7.0	94.0	97.0	-3
2.50	8.0	96.0	96.0	0
2.60	11.0	99.0	98.0	0

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=55.593 , m_C=59.889 , the average of B and C respectively.

Table 16, Comparision of cauchy distribution with seed=7118863

sample size n=5

Location shift δ	A	B	C	B-C
.00	4.0	1.0	16.0	-15
.10	1.0	.0	19.0	-19
.20	3.0	.0	13.0	-12
.30	1.0	1.0	13.0	-9
.40	3.0	.0	9.0	-14
.50	2.0	.0	14.0	-23
.60	3.0	2.0	25.0	-17
.70	2.0	1.0	18.0	-15
.80	4.0	2.0	17.0	-17
.90	3.0	6.0	23.0	-21
1.00	1.0	7.0	28.0	-21
1.10	1.0	6.0	27.0	-24
1.20	4.0	9.0	33.0	-16
1.30	3.0	12.0	28.0	-26
1.40	6.0	14.0	40.0	-24
1.50	4.0	12.0	36.0	-22
1.60	5.0	16.0	38.0	-31
1.70	3.0	18.0	49.0	-32
1.80	3.0	17.0	49.0	-31
1.90	5.0	31.0	62.0	-37
2.00	2.0	26.0	63.0	-35
2.10	5.0	31.0	66.0	-35
2.20	6.0	38.0	71.0	-33
2.30	4.0	44.0	74.0	-30
2.40	2.0	40.0	80.0	-40
2.50	2.0	46.0	76.0	-30
2.60	1.0	52.0	81.0	-29

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=16.0 , m_C=39.556 , the average of B and C respectively.

Table 17, Comparision of cauchy distribution with seed=7118863

sample size n=10

Location shift δ	A	B	C	B-C
.00	4.0	.0	14.0	-14
.10	4.0	.0	14.0	-14
.20	3.0	.0	10.0	-10
.30	7.0	2.0	19.0	-17
.40	5.0	2.0	14.0	-12
.50	4.0	5.0	15.0	-10
.60	1.0	5.0	17.0	-12
.70	3.0	4.0	22.0	-18
.80	8.0	8.0	26.0	-18
.90	6.0	10.0	23.0	-13
1.00	5.0	13.0	32.0	-21
1.10	6.0	12.0	28.0	-16
1.20	6.0	19.0	41.0	-22
1.30	3.0	30.0	58.0	-28
1.40	3.0	37.0	62.0	-25
1.50	6.0	39.0	66.0	-27
1.60	3.0	39.0	68.0	-29
1.70	9.0	44.0	69.0	-25
1.80	6.0	53.0	76.0	-23
1.90	2.0	52.0	81.0	-29
2.00	3.0	56.0	82.0	-26
2.10	1.0	63.0	92.0	-29
2.20	4.0	64.0	87.0	-23
2.30	.0	72.0	92.0	-20
2.40	6.0	78.0	93.0	-15
2.50	7.0	79.0	96.0	-17
2.60	3.0	83.0	97.0	-14

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=32.185 , m_C=51.630 , the average of B and C respectively.

Table 18, Comparision of cauchy distribution with seed=7118863

sample size n=15

Location shift δ	A	B	C	B-C
.00	2.0	.0	12.0	-12
.10	6.0	.0	15.0	-15
.20	6.0	1.0	13.0	-12
.30	5.0	.0	8.0	-8
.40	5.0	2.0	14.0	-12
.50	5.0	2.0	17.0	-15
.60	5.0	9.0	19.0	-10
.70	7.0	13.0	34.0	-21
.80	4.0	20.0	32.0	-12
.90	6.0	17.0	29.0	-12
1.00	8.0	26.0	46.0	-20
1.10	11.0	24.0	51.0	-27
1.20	9.0	44.0	62.0	-18
1.30	8.0	36.0	63.0	-27
1.40	2.0	52.0	77.0	-25
1.50	5.0	64.0	85.0	-21
1.60	4.0	62.0	83.0	-21
1.70	8.0	77.0	92.0	-15
1.80	2.0	74.0	89.0	-15
1.90	4.0	79.0	90.0	-11
2.00	2.0	87.0	95.0	-8
2.10	3.0	85.0	96.0	-11
2.20	7.0	86.0	96.0	-10
2.30	5.0	94.0	97.0	-3
2.40	4.0	91.0	97.0	-6
2.50	7.0	91.0	98.0	-7
2.60	3.0	93.0	100.0	-7

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=45.519 , m_C=59.630 , the average of B and C respectively.

Table 19, Comparision of cauchy distribution with seed=7118863

sample size n=20

Location shift δ	A	B	C	B-C
.00	8.0	2.0	11.0	-9
.10	6.0	.0	9.0	-9
.20	9.0	2.0	10.0	-8
.30	3.0	4.0	11.0	-7
.40	8.0	4.0	18.0	-14
.50	5.0	5.0	12.0	-7
.60	10.0	16.0	26.0	-10
.70	3.0	18.0	29.0	-11
.80	7.0	22.0	44.0	-22
.90	8.0	32.0	47.0	-15
1.00	5.0	36.0	54.0	-18
1.10	3.0	47.0	65.0	-18
1.20	3.0	60.0	72.0	-12
1.30	9.0	71.0	82.0	-11
1.40	5.0	68.0	78.0	-10
1.50	6.0	81.0	90.0	-9
1.60	3.0	82.0	93.0	-11
1.70	7.0	89.0	93.0	-4
1.80	3.0	90.0	96.0	-6
1.90	5.0	94.0	97.0	-3
2.00	5.0	95.0	99.0	-4
2.10	6.0	96.0	97.0	-1
2.20	9.0	97.0	97.0	0
2.30	7.0	99.0	100.0	-1
2.40	9.0	96.0	100.0	-4
2.50	7.0	99.0	99.0	0
2.60	7.0	98.0	100.0	-2

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=55.67 , m_C=64.04 , the average of B and C respectively.

Table 20, Comparision of cauchy distribution with seed=7118863

sample size n=25

Location shift δ	A	B	C	B-C
.00	5.0	1.0	8.0	-7
.10	7.0	2.0	10.0	-8
.20	5.0	4.0	19.0	-15
.30	4.0	3.0	11.0	-8
.40	7.0	8.0	16.0	-8
.50	5.0	20.0	30.0	-10
.60	8.0	19.0	25.0	-6
.70	14.0	26.0	35.0	-9
.80	3.0	35.0	48.0	-13
.90	8.0	42.0	58.0	-16
1.00	5.0	56.0	60.0	-4
1.10	8.0	68.0	77.0	-9
1.20	5.0	72.0	80.0	-8
1.30	10.0	75.0	82.0	-7
1.40	5.0	83.0	90.0	-7
1.50	10.0	90.0	93.0	-3
1.60	7.0	95.0	95.0	0
1.70	9.0	96.0	98.0	-2
1.80	8.0	96.0	97.0	-1
1.90	10.0	99.0	99.0	0
2.00	7.0	98.0	100.0	-2
2.10	4.0	100.0	100.0	0
2.20	12.0	99.0	99.0	0
2.30	11.0	100.0	100.0	0
2.40	8.0	100.0	100.0	0
2.50	3.0	100.0	100.0	0
2.60	13.0	99.0	99.0	0

A:power of $\bar{x}$ (No. of rejections in 100 samples)
B:power of Mann-Whitney
C:power of Huber
m_B=62.44 , m_C=67.74 , the average of B and C respectively.

Chapter 7

RESULTS AND CONCLUSION

The curve ot operating characteristic is used to judge the effectiveness of three cases: (1) use the standard normal estimate. (2) use Mann- Whitney U test (nonparametric method). (3) Huber estimate of the location and scale.

In many settings, we have only a vague notion of the distributon generating the data. What is needed,then ,is not an estimator that is best for a specific situation , but rather one that is rairly good over a suitable range of situations. We want an estimator that has fairly high erficiency over a range of possible situations. And the advantage of any robust estimator in routine use depends more on its erficiency in small samples than on its optimality in large samples.

In this study , we consider sample size 5,10,15 and 20,25 on three cases . Figures from 2 to 6 show comparison of contaminated distribution when H=3 p=.05 under three different conditions and sample size 5,10,15,20 and 25 respectively. In figure 2, even the sample size is as small as 5 , the Huber estimator make the control the best except when there is no shift, the Mann-Whitney U test make the control the worst one. So Mann-Whitney U test (nonparametric method) is not good when sample size is too small. When sample size is greater or equal 10, the Huber estimator still be the best, and Mann-Whitney U test is better than the standard normal estimator. So Huber estimator is considered to be the robust estimator under the case of contaminated normal condition.

Figures from 7 to 11 show comparison of slash distribution under three different conditions and sample size 5,10,15,20 and 25 respectively. In figure 7 , when the sample size is as small as 5, and the location shift is less than 1.5, the Mann-Whitney U test is the same bad as the standard normal estimate. We knew that we can not find the expected value of mean under slash distribution, so standard normal estimate do not work under slash distribution even the sample size is larger. The Huber estimator make the control the best except when there is no shift. When sample size is greater or equal 10, the Mann-Whitney U test is better, and the Huber estimator is the best. So Huber estimator is considered to be the robust estimator under the case of slash distribution. Figures from 12 to 16 show comparison of cauchy distribution under three different conditions and sample size 5, 10,15,20 and 25 respectively. We knew that we can not find the expected value of mean under cauchy distribution, so standard normal estimate do not work under cauchy distribution either. In figure 12, When sample size is as small as 5, and the location shift is less than 2.5, the Mann-Whitney U test is the same bad as the standard normal estimate. The Huber estimator make the control the best except when there is no shift. When sample size is greater or equal 10 , the Mann-Whitney U test is better, and the Huber estimator is still be the best. So Huber estimator is the robust estimator under the case of cauchy distribution.

From table 6 to 20, we can use t-statistic to test the difference between $\bar{x}$ and Mann-Whitney , $\bar{x}$ and Huber , Mann-Whitney and Huber.

$$t = \frac{m_x - m_y}{S_{x-y}/n^{1/2}} \qquad \text{with (n-1) degrees of freedom}$$

Here n is n pairs of reading, S_{x-y} stands for the standard deviation of paired difference (x-y), i.e.

$$S_{x-y} = \sqrt{\frac{\sum[(x-y)-(m_x-m_y)]^2}{n-1}}$$

From table 6, $S_{A-B}=3.125$, $t_{A-B}=3.52>2.131$, with 5 degrees of freedom, it falls almost exactly at 5% level of significance. $\bar{x}$ and MW appears to have difference. $S_{A-C}=4.61$, $t_{A-C}=4.176 > 2.131$, $S_{B-C}=10.18$ $t_{B-C}=2.96 > 2.131$, so, it appears to have difference among $\bar{x}$, MW, and Huber.

From table 7, $S_{A-B}=3.48$, $t_{A-B}=1.797<2.131$, there is no significant difference between $\bar{x}$ and MW at 5% level of significance. $S_{A-C}=3.39$, $t_{A-C}=4.94>2.131$, $\bar{x}$ and Huber have difference. $S_{B-C}=3.03$,

$t_{B-C}=3.46>2.131$, MW and Huber have difference, with 15 degrees of freedom.

From table 8, $S_{A-B}=4.22$, $t_{A-B}=3.674>2.131$, $S_{A-C}=3.76$, $t_{A-C}=4.66>2.131$, it appears to have difference between $\bar{x}$ and MW, and, between $\bar{x}$ and Huber. $S_{B-C}= 2.53$, $t_{B-C}=0.789<2.131$, it appears to have no difference between MW and Huber with 15 degrees of freedom.

From table 9, $S_{A-B}=3.51$, $t_{A-B}=4.3>2.131$, and $S_{A-C}=3.85$, $t_{A-C}=4.0>2.131$, it appears to have difference between $\bar{x}$ and MW, and, between $\bar{x}$ and Huber. $S_{B-C}=1.41$, $t_{B-C}=1.4<2.131$, it appears to have no difference MW and Huber with 15 degrees of freedom.

From table 10, $S_{A-B}=5.45$, $t_{A-B}=3.76>2.131$, $S_{A-C}=3.29$, $t_{A-C}=3.878>2.131$, it appears to have difference between $\bar{x}$ and MW, and, between $\bar{x}$ and Huber. $S_{B-C}=3.696$, $t_{B-C}=2.121<2.131$, it appears to have no difference between MW and Huber with 15 degrees of freedom.

From table 11 to table 20, there are significant difference between $\bar{x}$ and MW, and, $\bar{x}$ and Huber, so the test is not necessary. We only test the difference between MW and Huber. From table 11, $S_{B-C}=5.755$, $t_{B-C}=12.74>2.056$. From table 12, $S_{B-C}=5.137$, $t_{B-C}=11.83>2.056$. From table 13, $S_{B-C}=4.81$, $t_{B-C}=14.7>2.056$. From table 14, $S_{B-C}=3.29$, $t_{B-C}=11.99>2.056$. From table 15, $S_{B-C}=3.51$, $t_{B-C}=6.36>2.056$. From table 16, $S_{B-C}=8.31$, $t_{B-C}=14.72>2.056$. From table 17, $S_{B-C}=6.05$, $t_{B-C}=16.76>2.056$. From table 18, $S_{B-C}=6.06$, $t_{B-C}=12.06>2.056$. From table 19, $S_{B-C}=6.275$, $t_{B-C}=6.936>2.056$. From table 20, $S_{B-C}=4.9$, $t_{B-C}=5.62>2.056$. With 26 degrees of freedom and 5% level of significance, all these t values appear to have difference between MW and Huber.

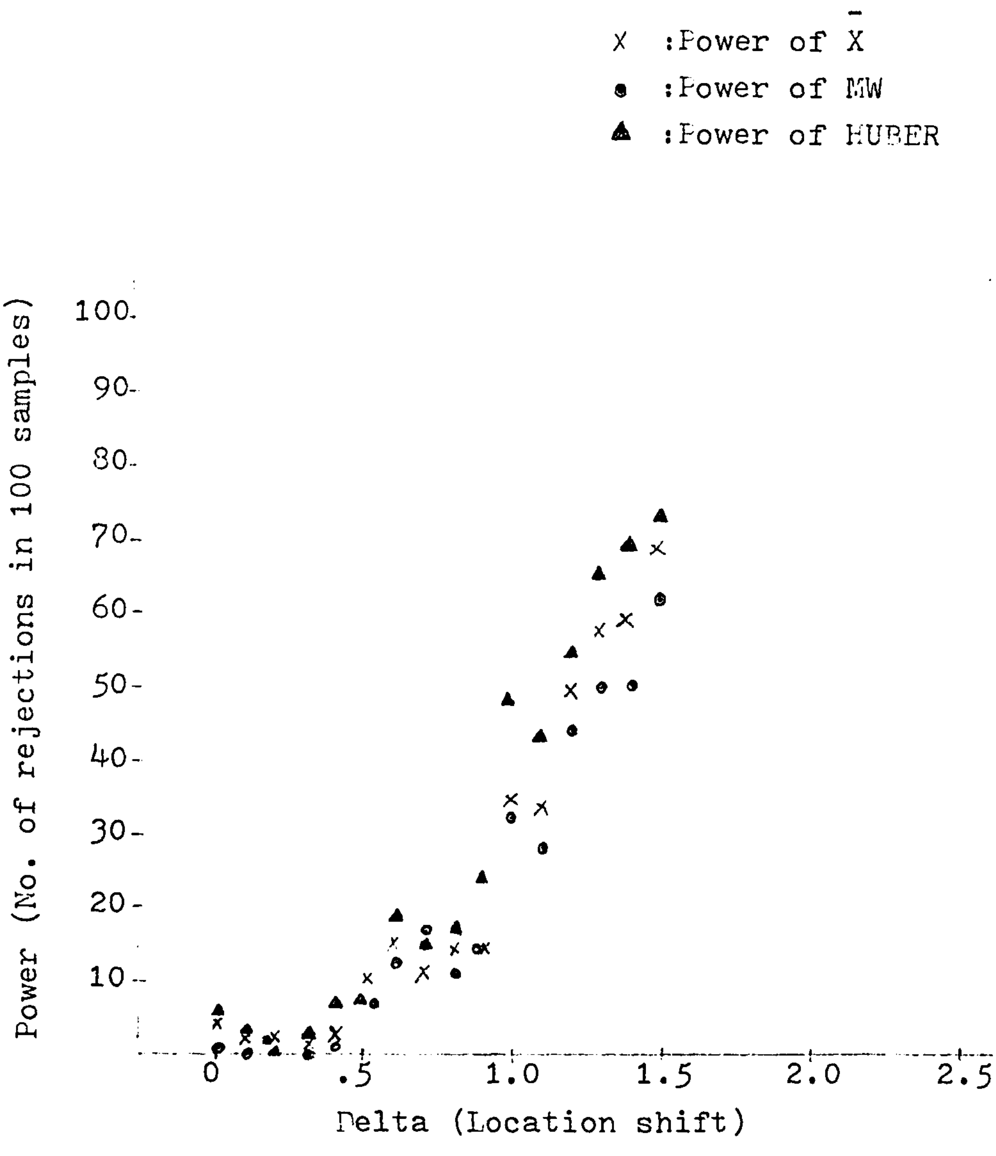

Fig.2, Compare of contaminated dist.
H=3 p=.05 seed=7118863, n=5

x :Power of $\bar{X}$
o :Power of MW
Δ :Power of HUBER

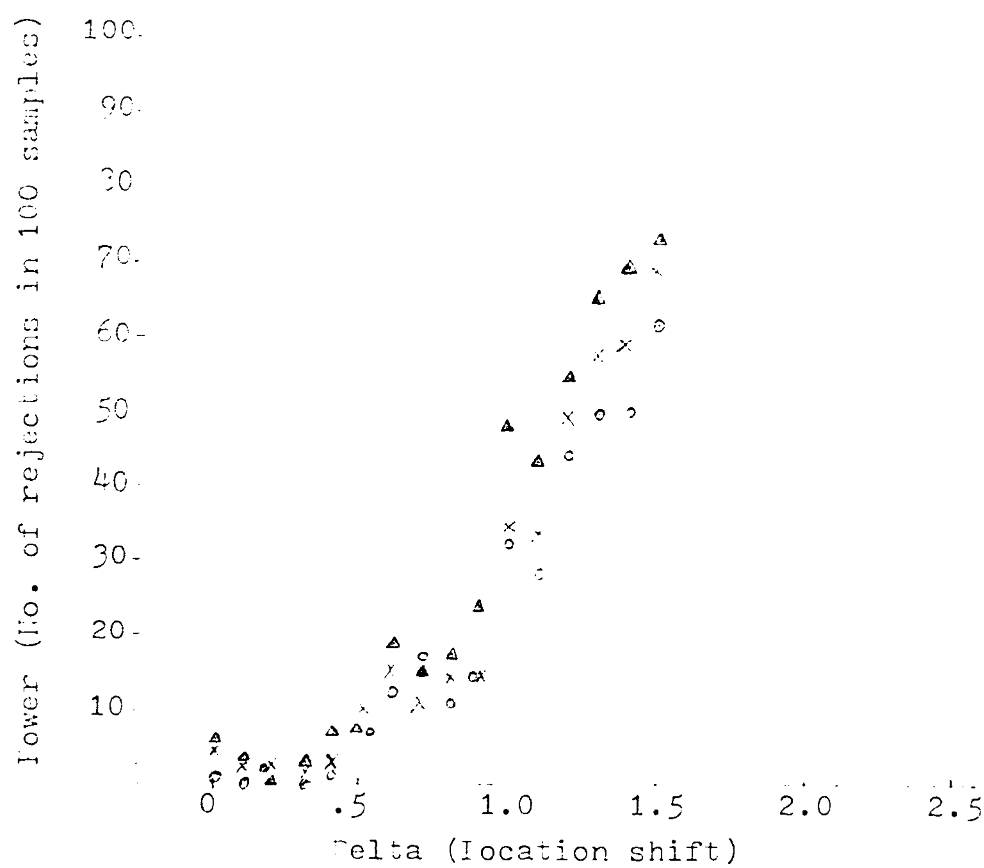

Fig.2, Compare of contaminated dist.
P=3 p=.05 seed=7118363, n=5

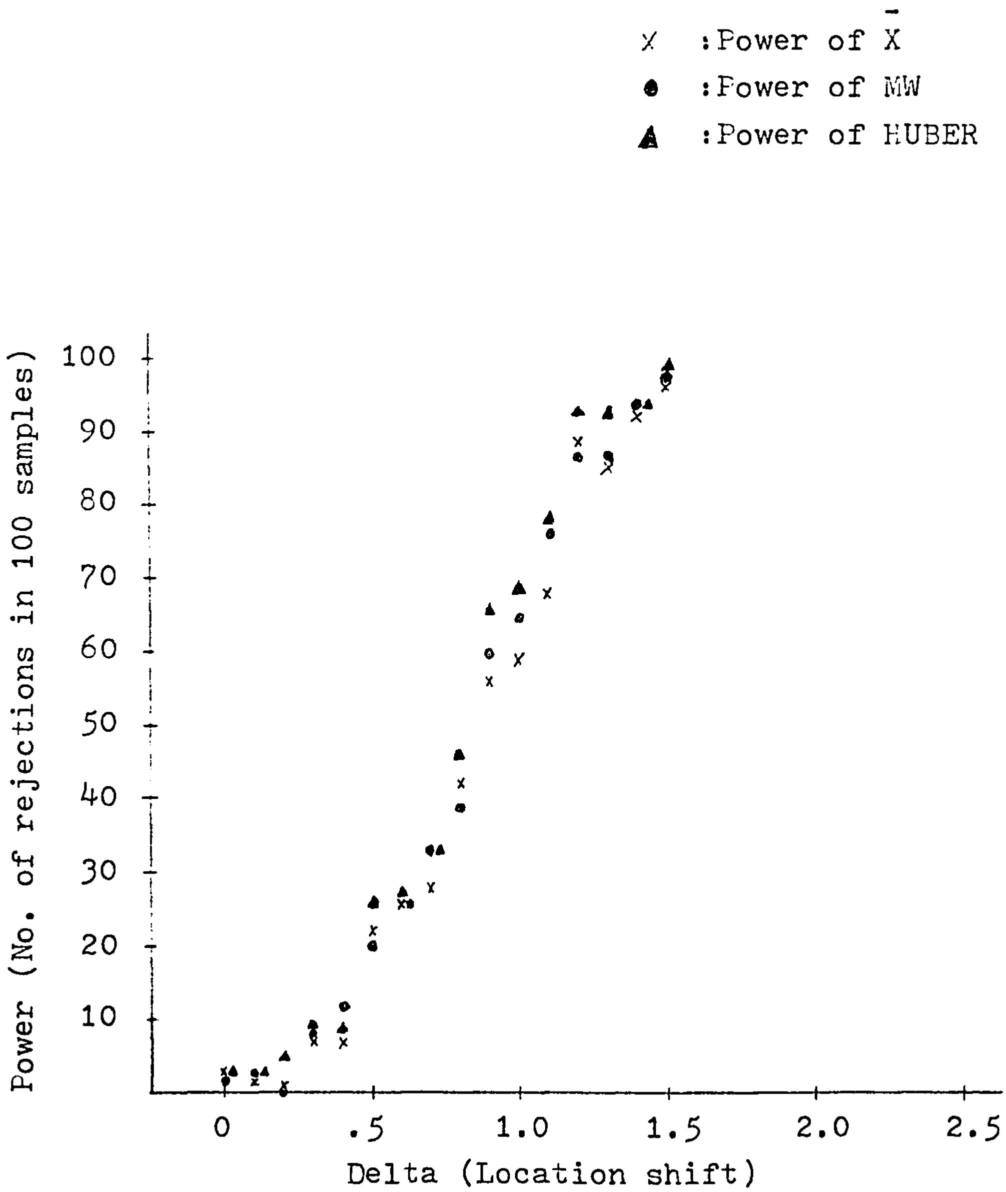

Fig.3, Compare of contaminated dist.
H=3 p=.05 seed=7118863, n=10

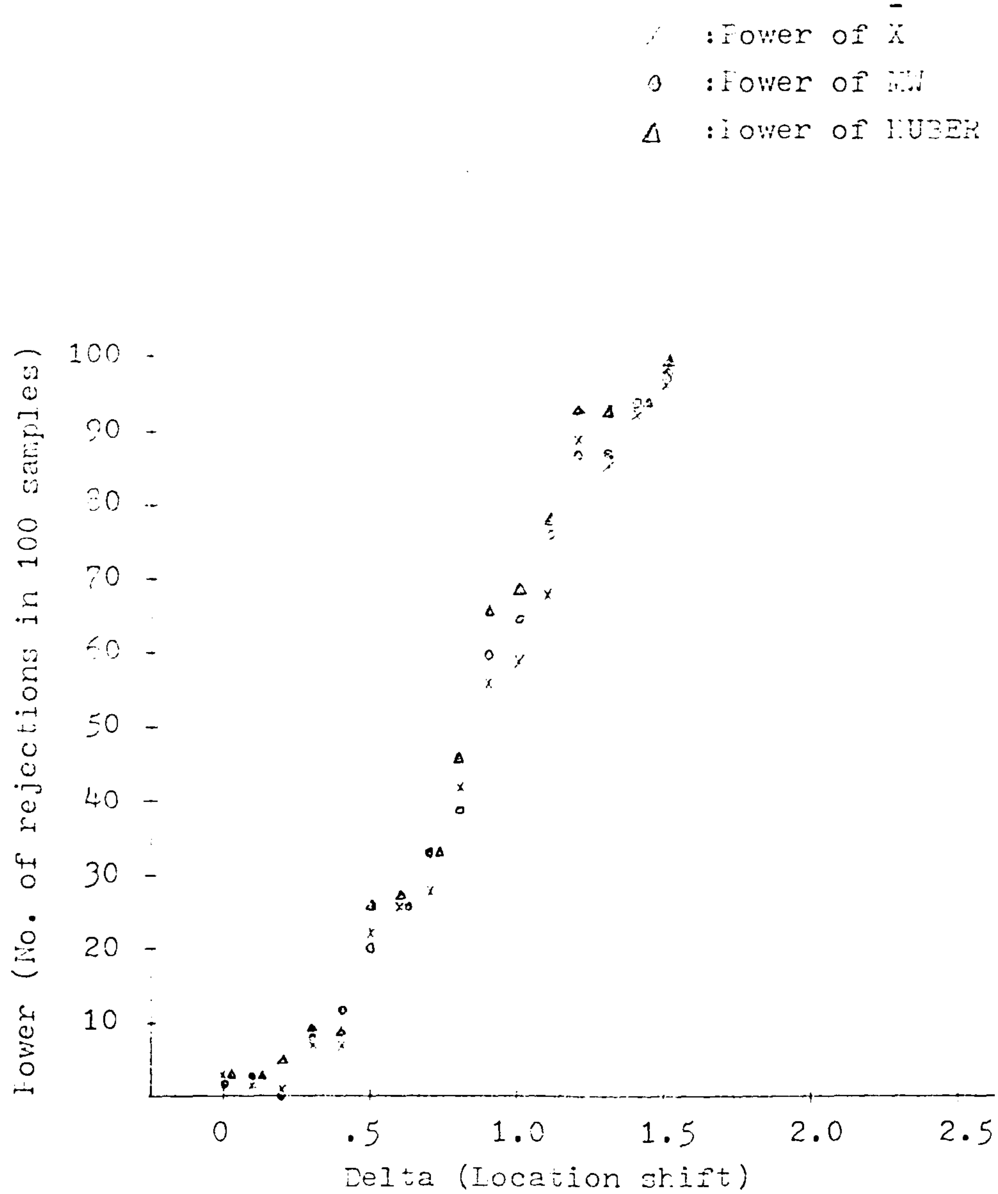

Fig.3, Compare of contaminated dist.
K=3 p=.05 seed=7118863, n=10

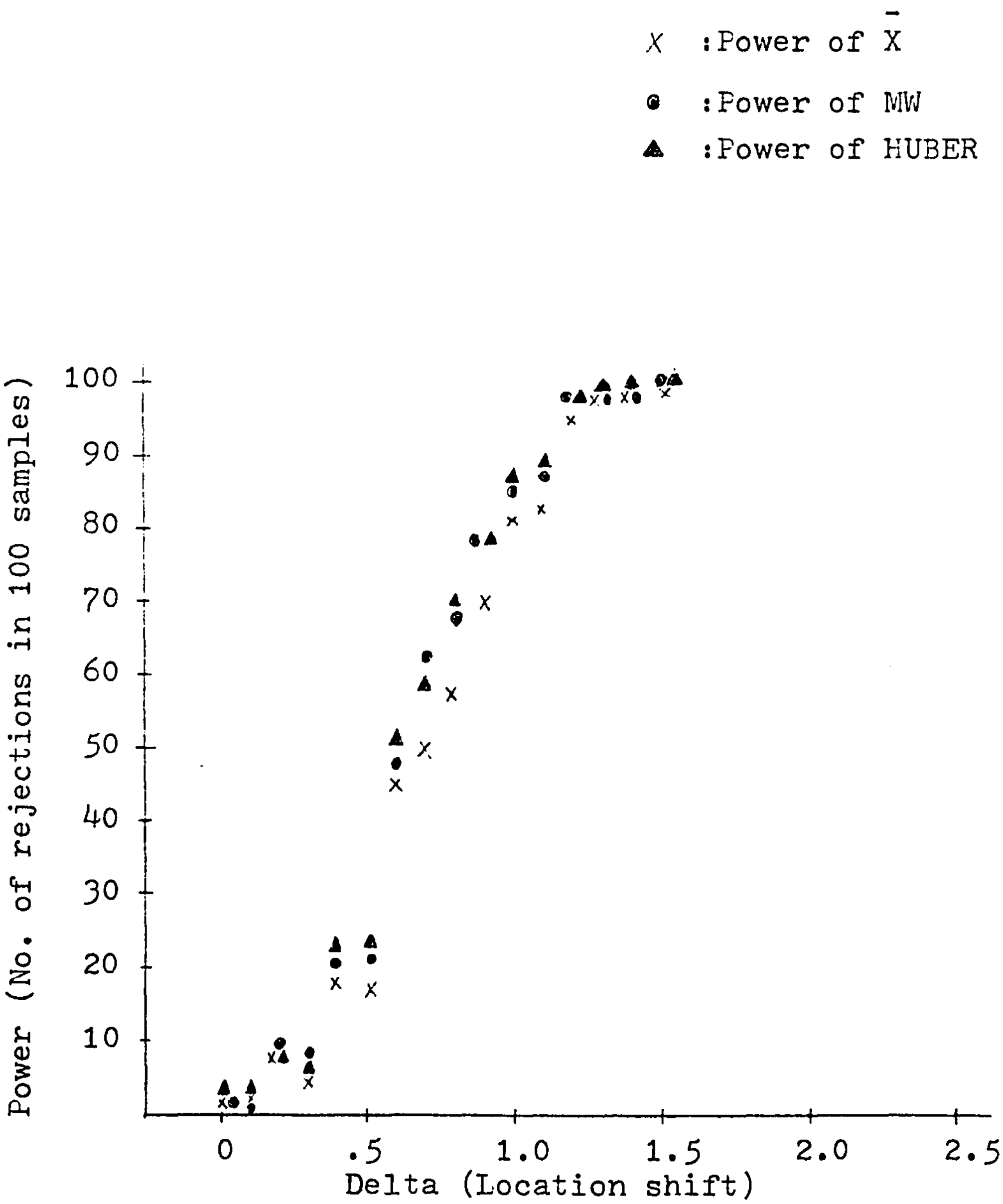

Fig.4, Compare of contaminated dist.
H=3 p=.05 seed=7118863, n=15

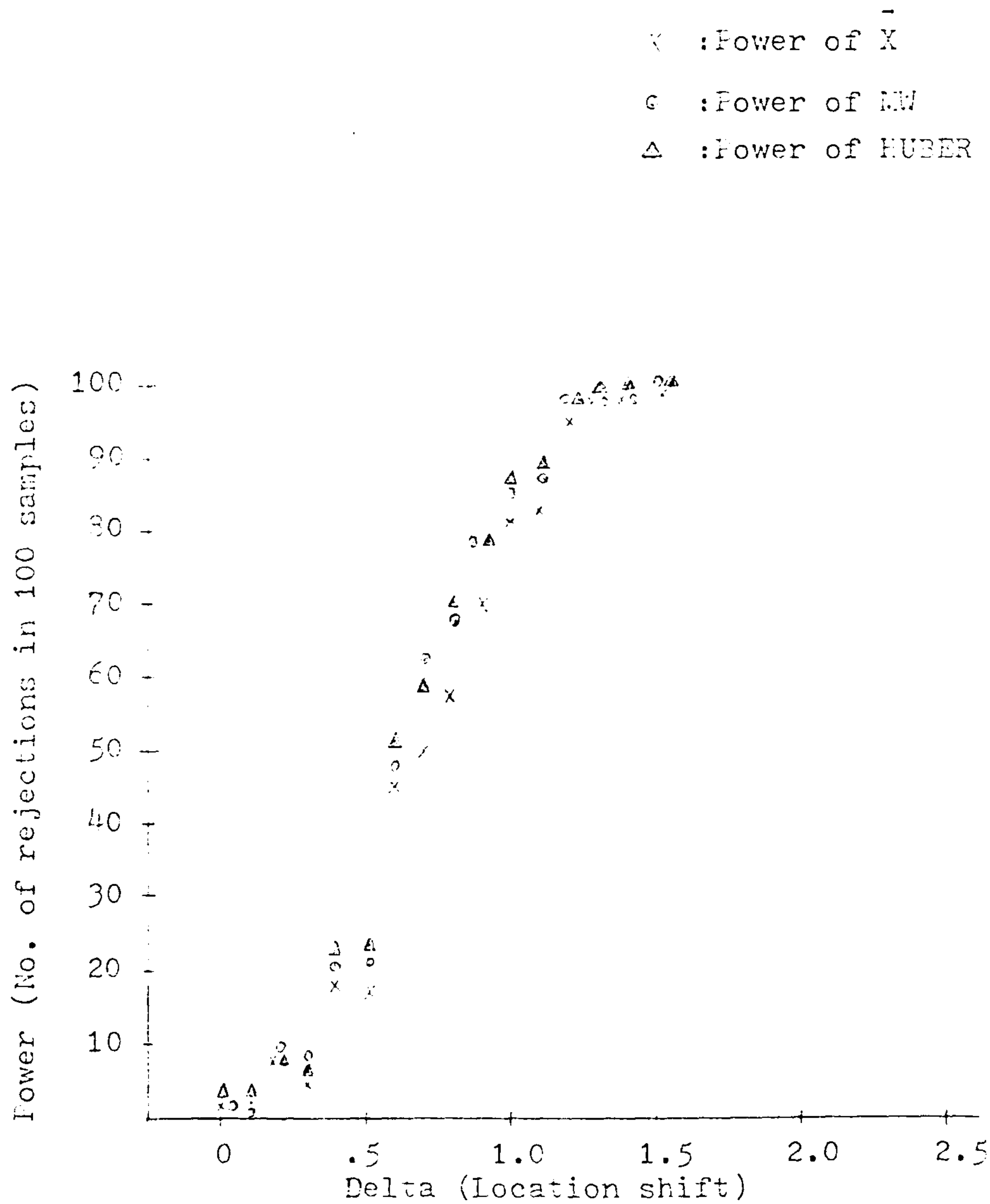

Fig.4, Compare of contaminated dist.
k=3 p=.05 seed=7118863, n=15

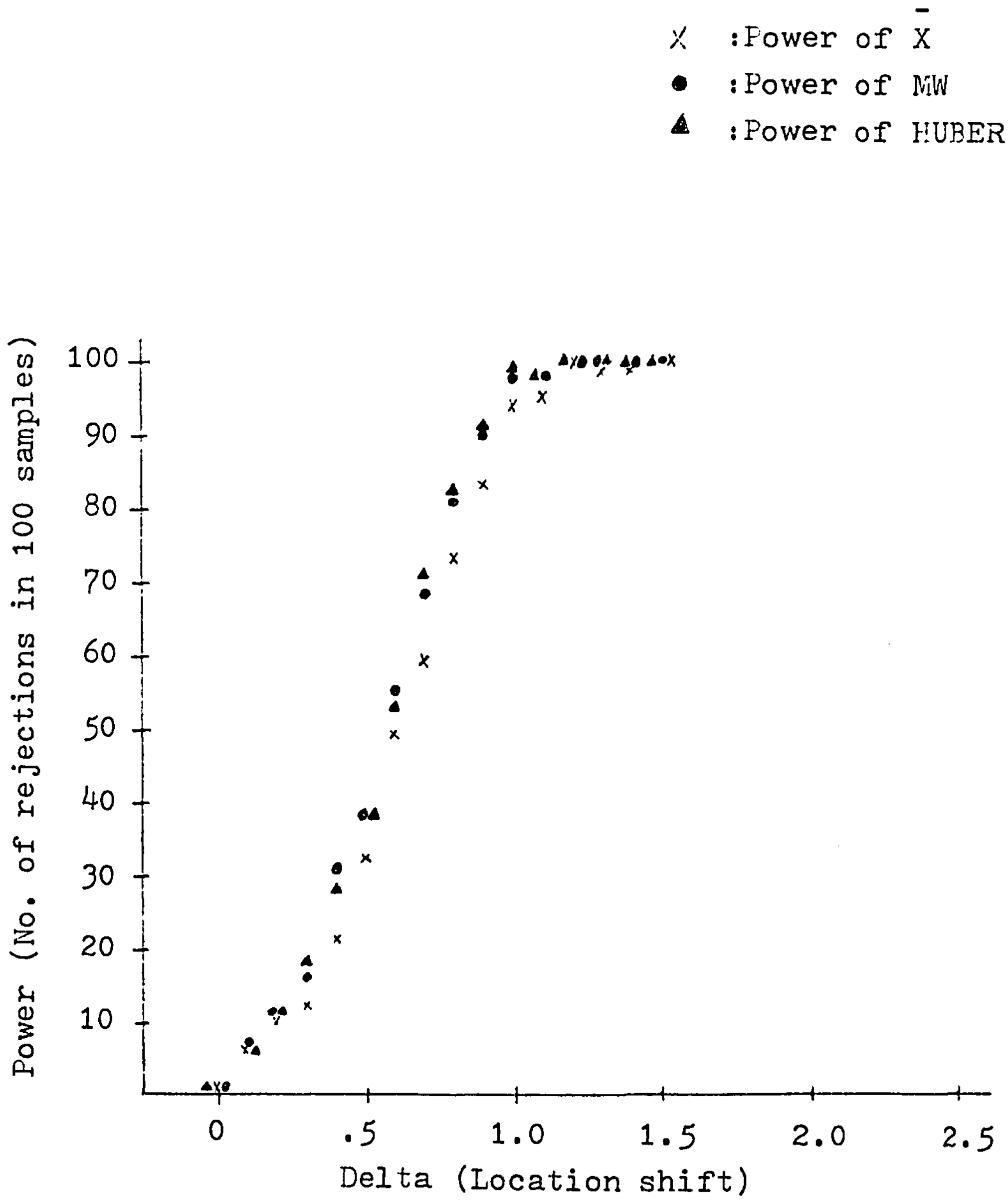

Fig.5, Compare of contaminated dist.
H=3 p=.05 seed=7118863, n=20

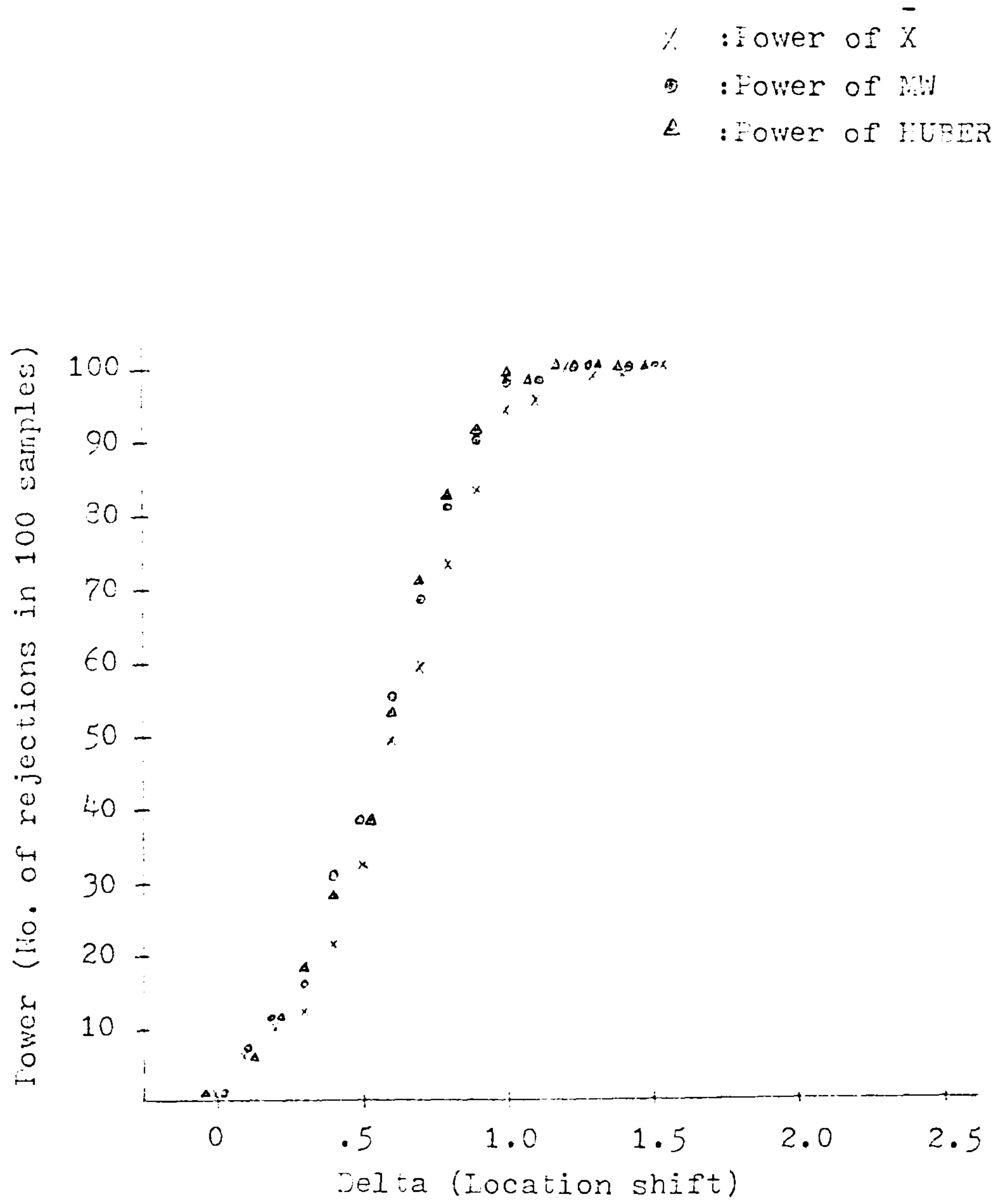

Fig.5, Compare of contaminated dist.
H=3 p=.05 seed=7118863, n=20

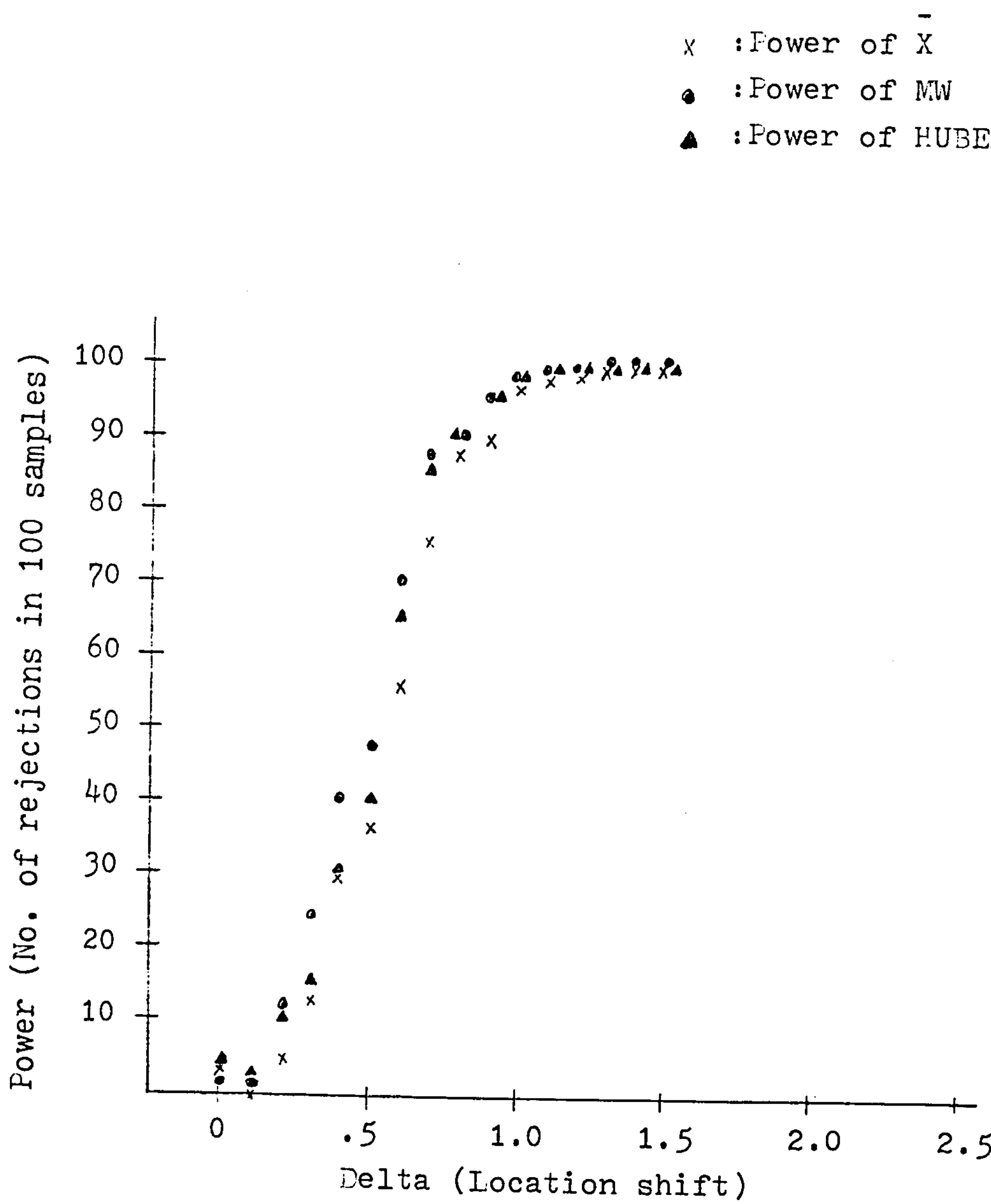

Fig.6, Compare of contaminated dist.
H=3 p=.05 seed=7118863, n=25

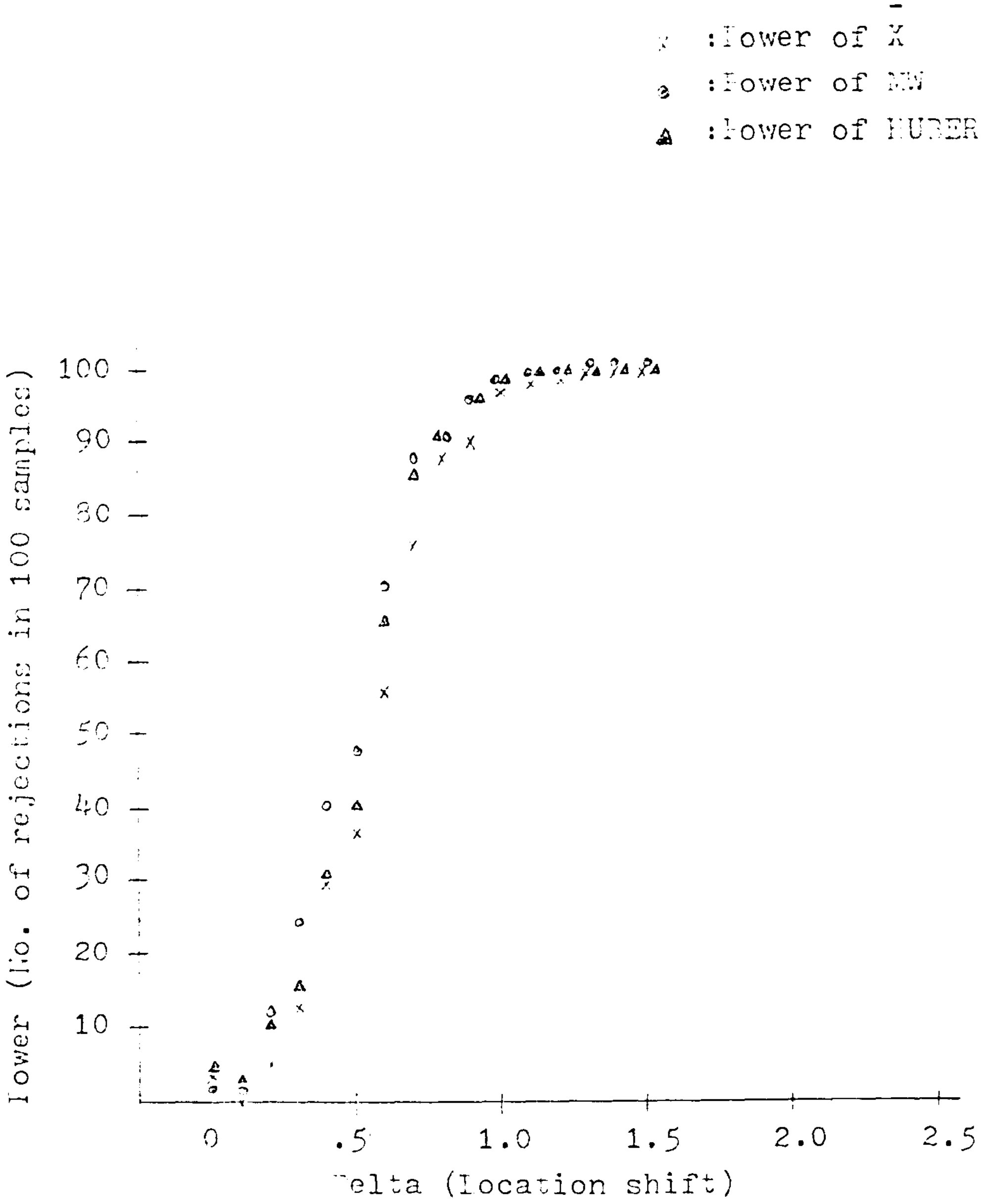

Fig.6, Compare of contaminated dist.
K=3 p=.05 seed=7113863, n=25

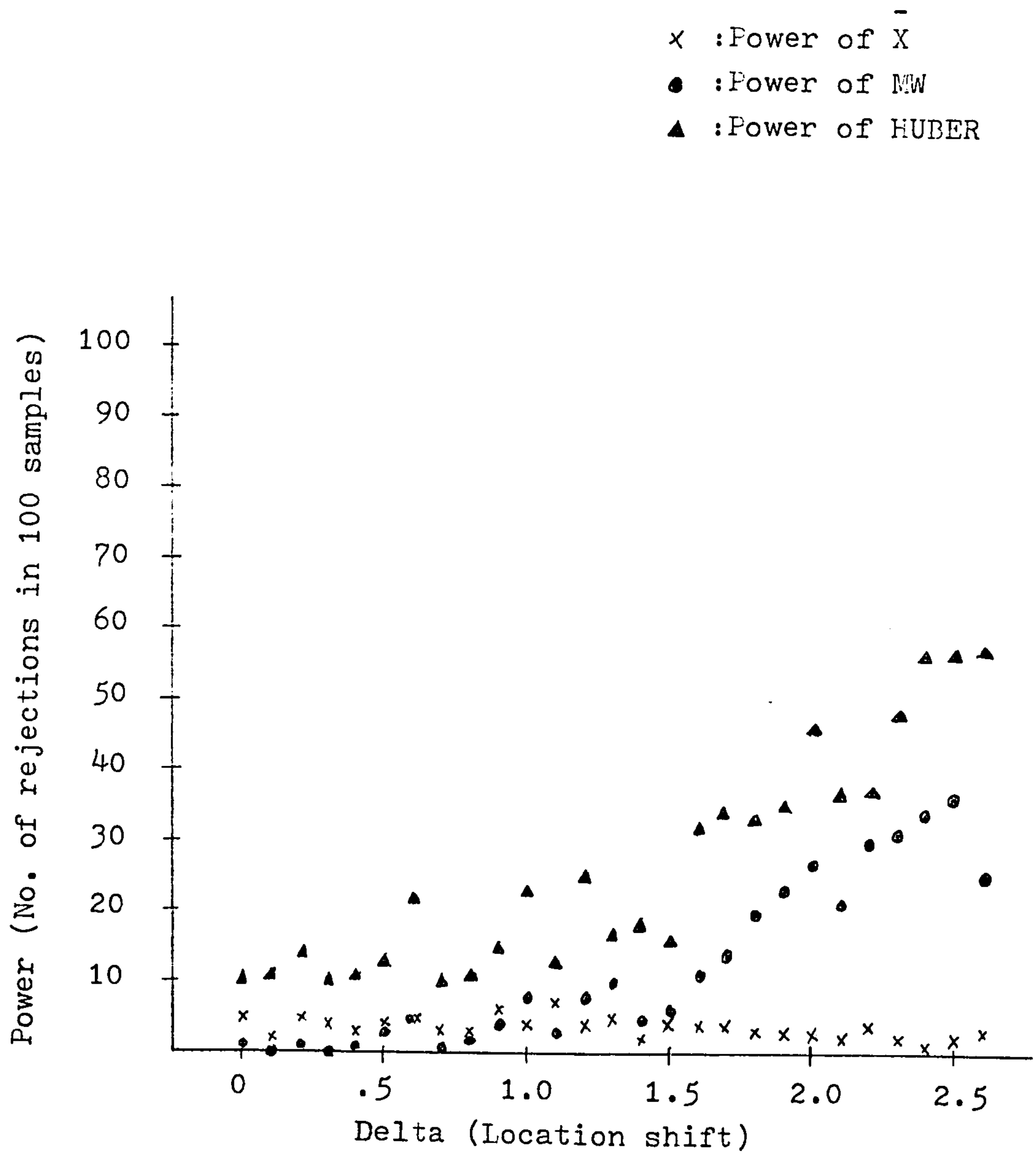

Fig.7, Compare of slash dist.
seed=7118863, n=5

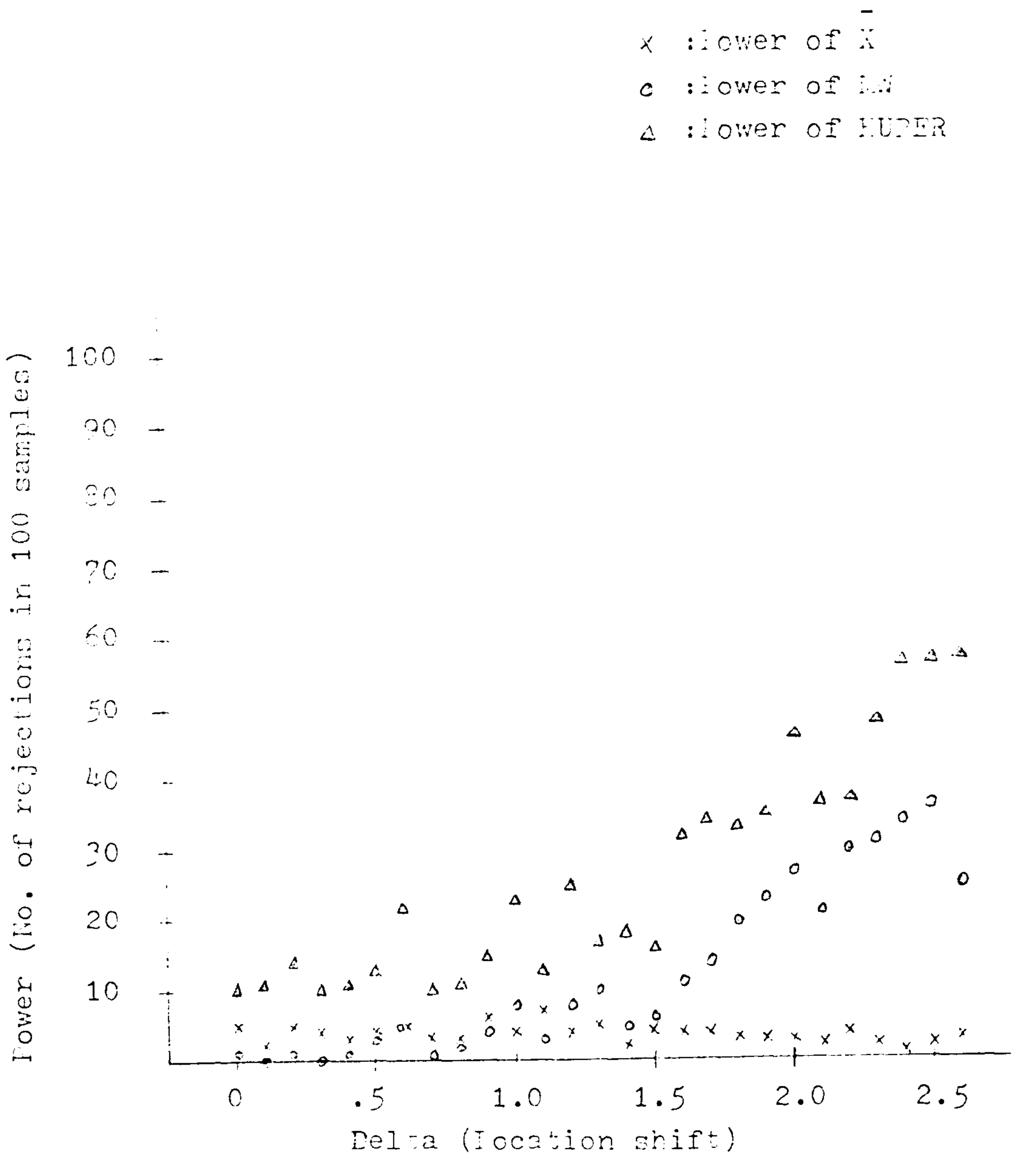

Fig.7, Compare of slash dist.
seed=7118863, n=5

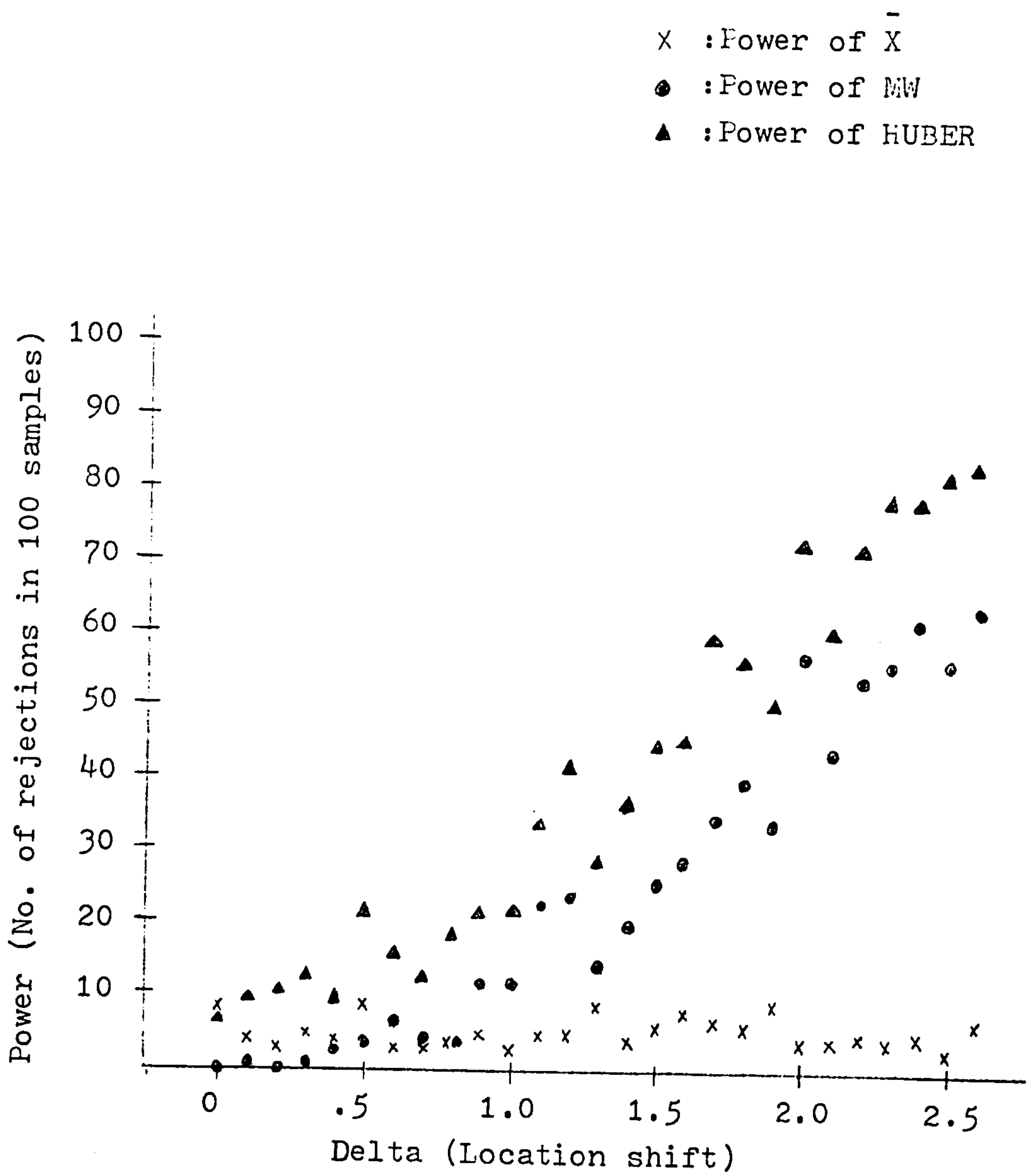

Fig.8, Compare of slash dist.
seed=7118863, n=10

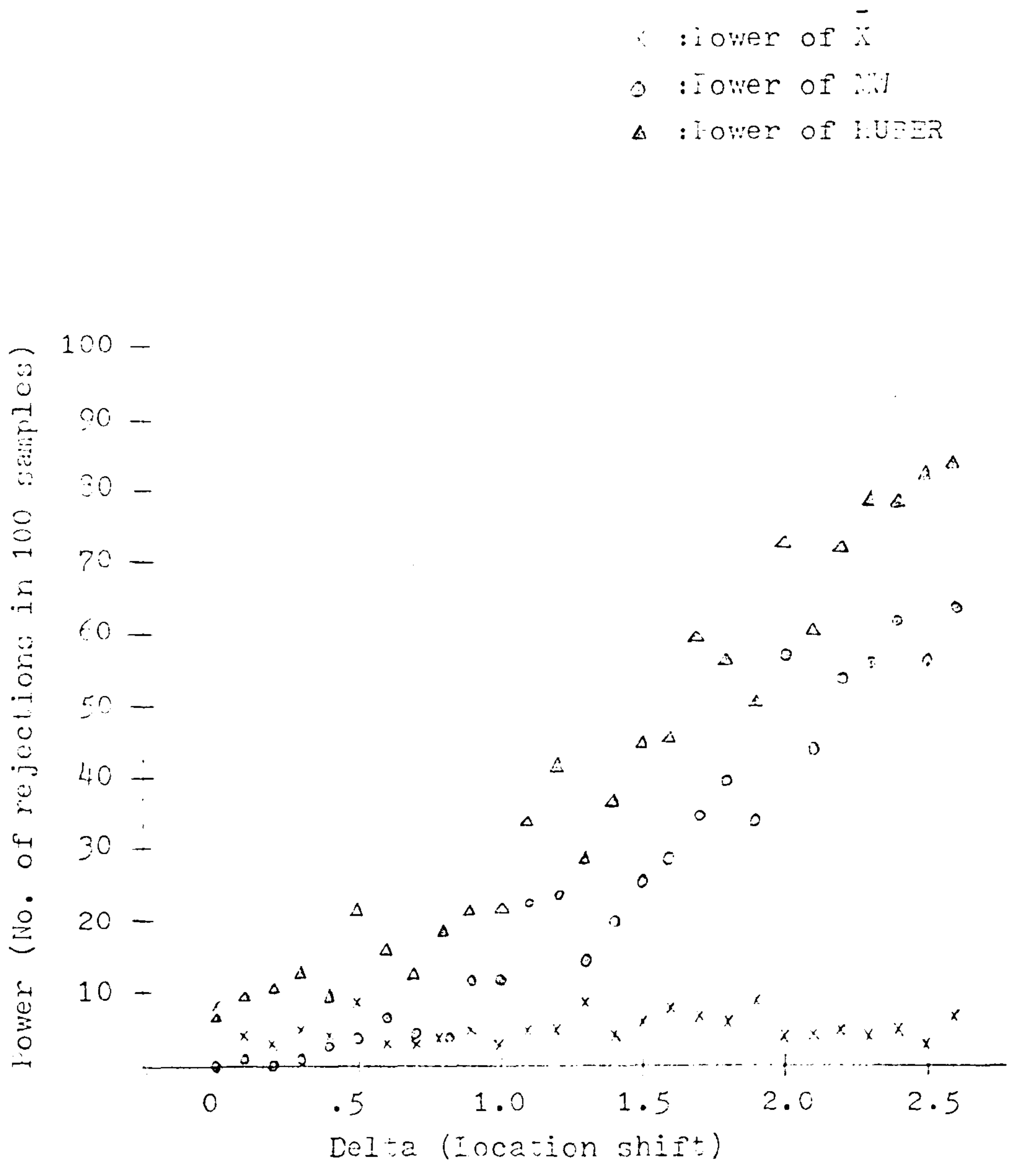

Fig.8, Compare of slash dist.
seed=7113863, n=10

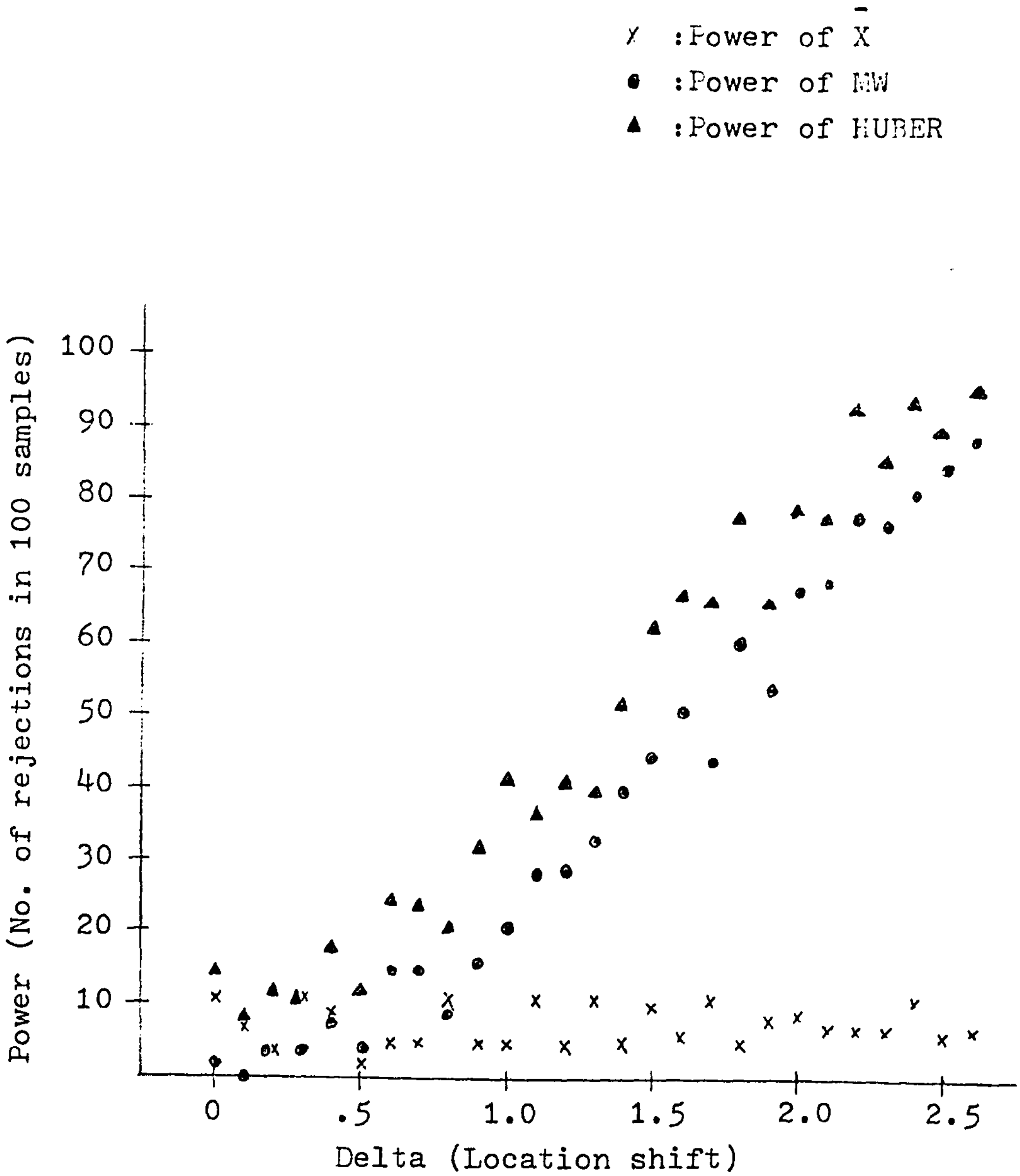

Fig.9, Compare of slash dist.
seed=7118863, n=15

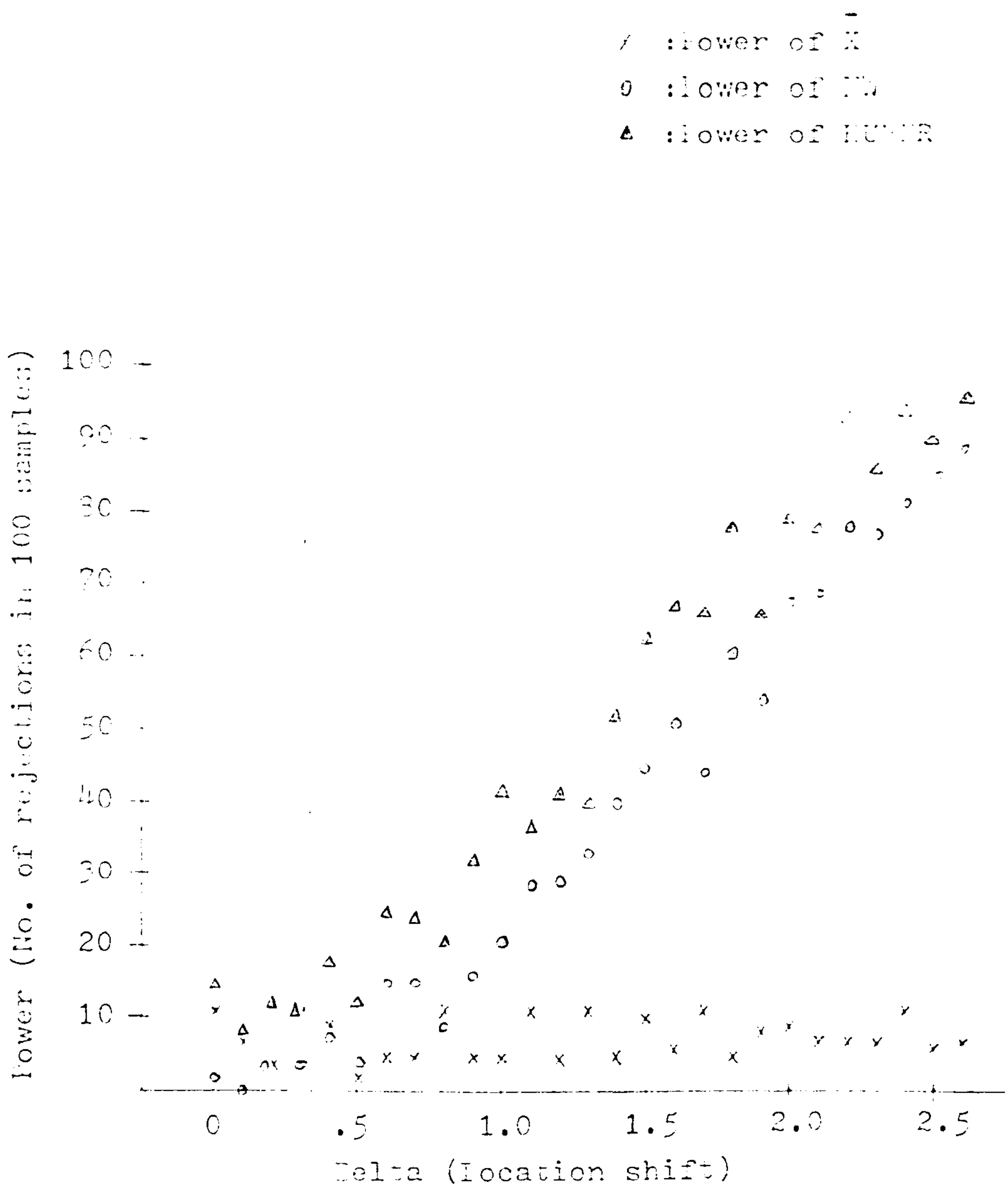

Fig.9, Compare of slash dist.
seed=7118863, n=15

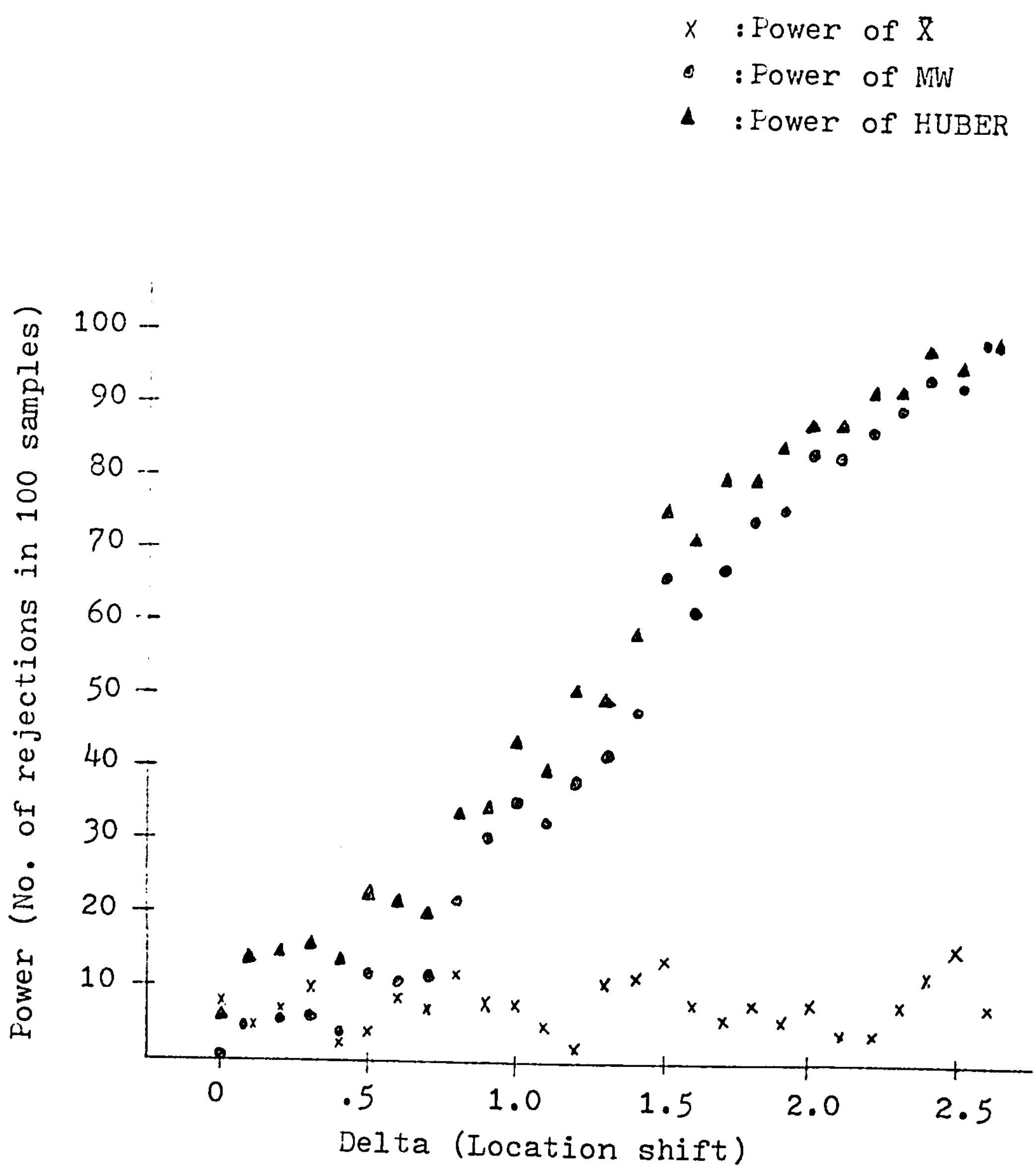

Fig.10, Compare of slash dist.
seed=7118863, n=20

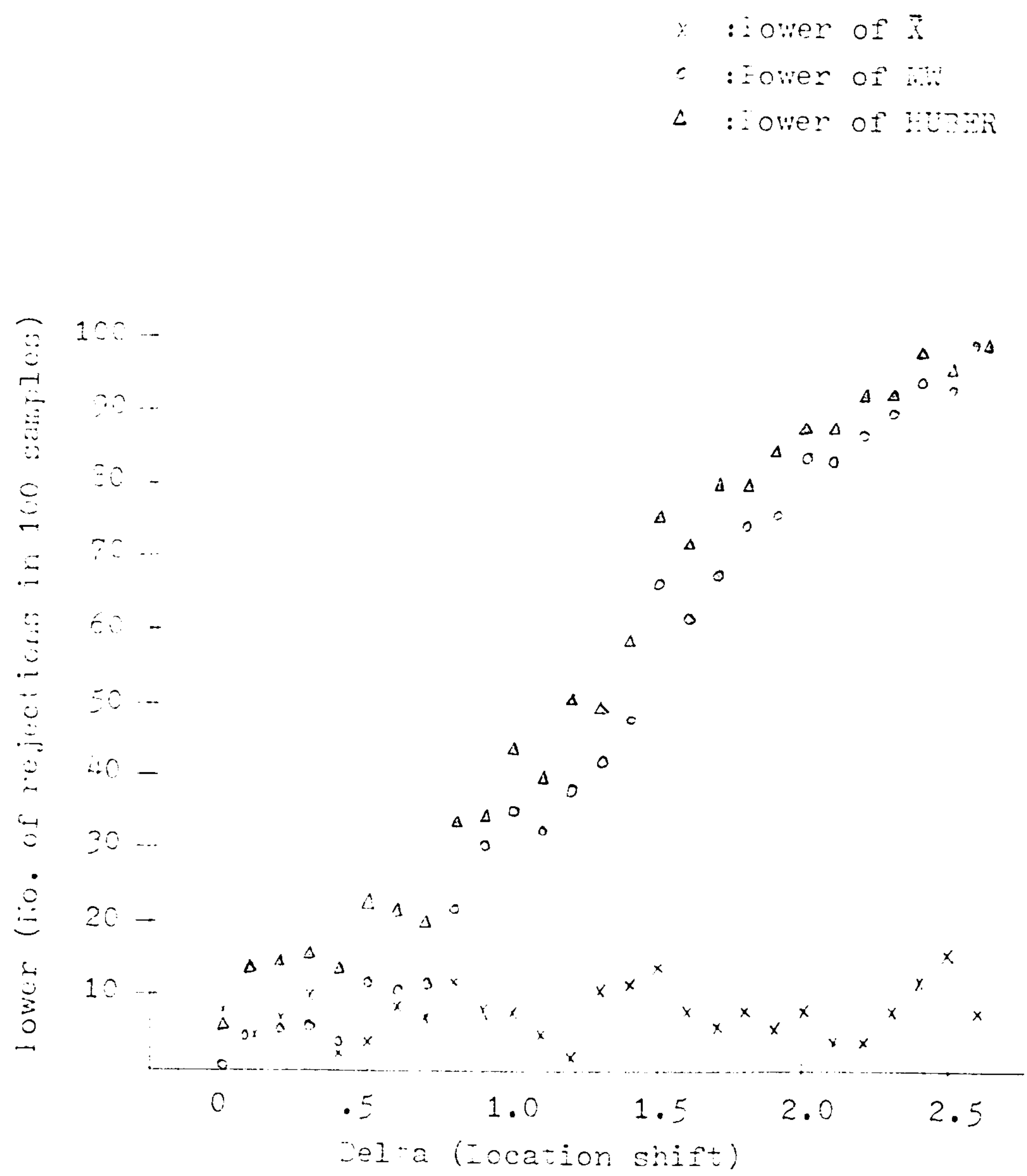

Fig.10, Compare of slash dist.
seed=7113863, n=20

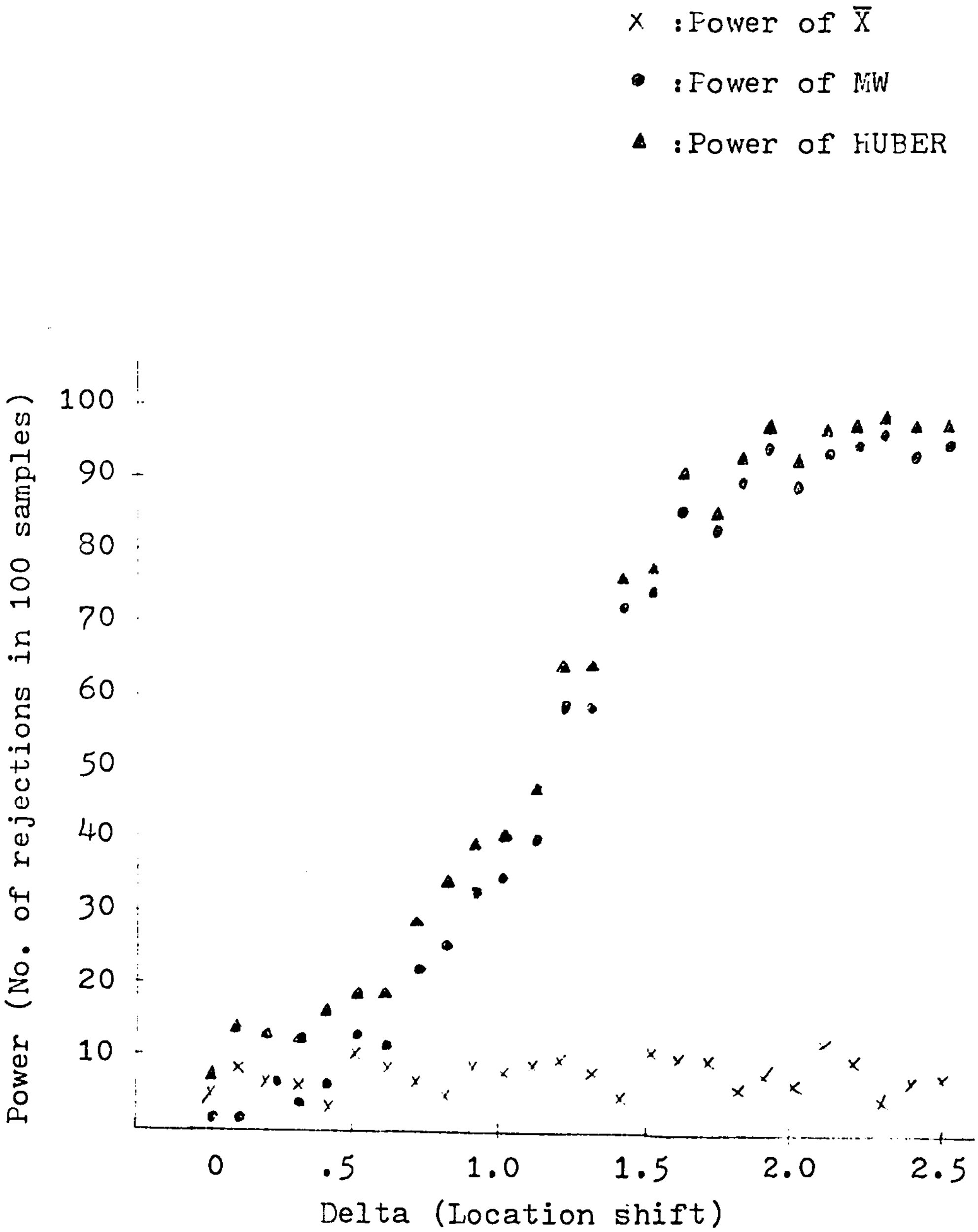

Fig.11, Compare of slash dist.
seed=7118863, n=25

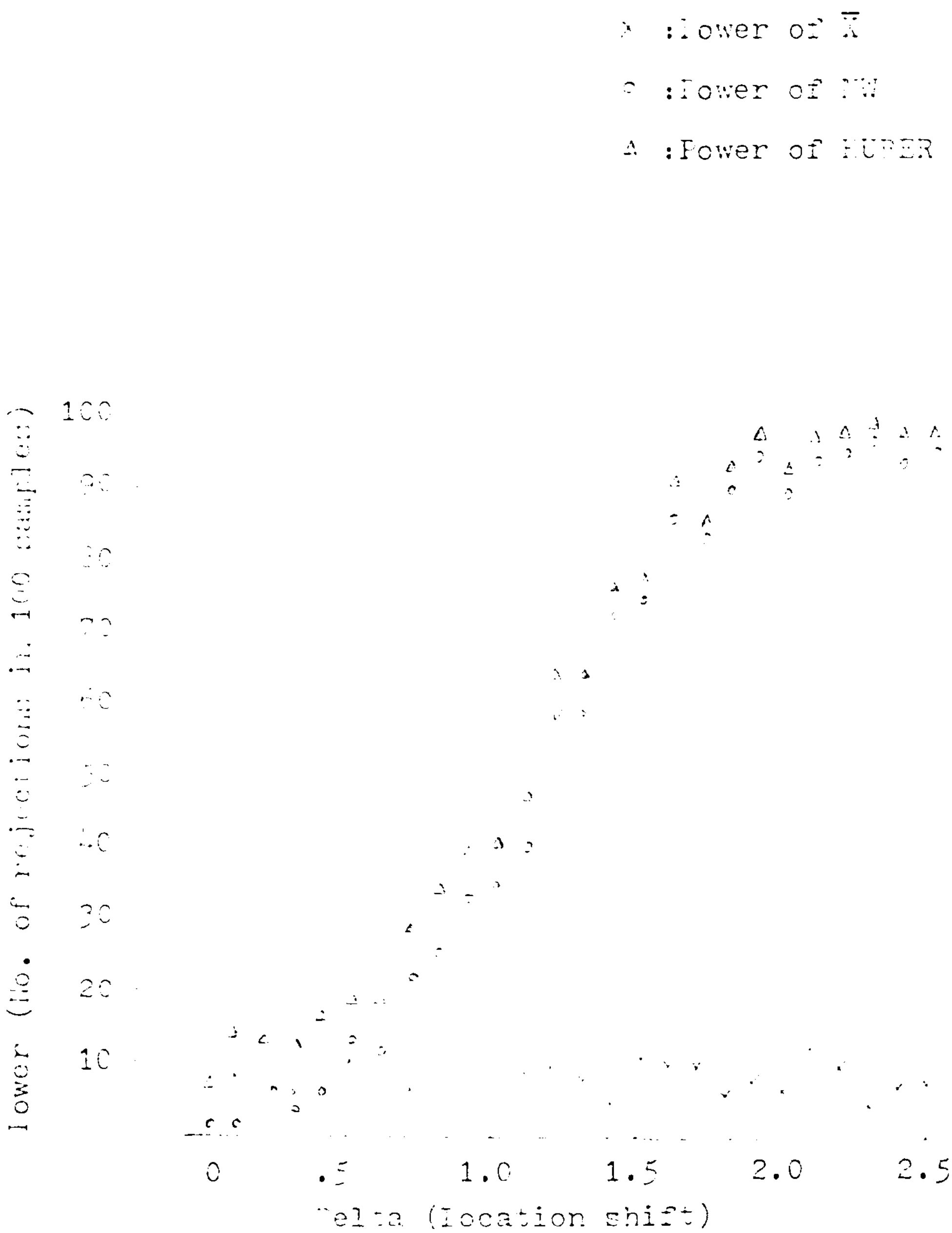

Fig.11, Compare of Slash dist.
seed=7113363, n=25

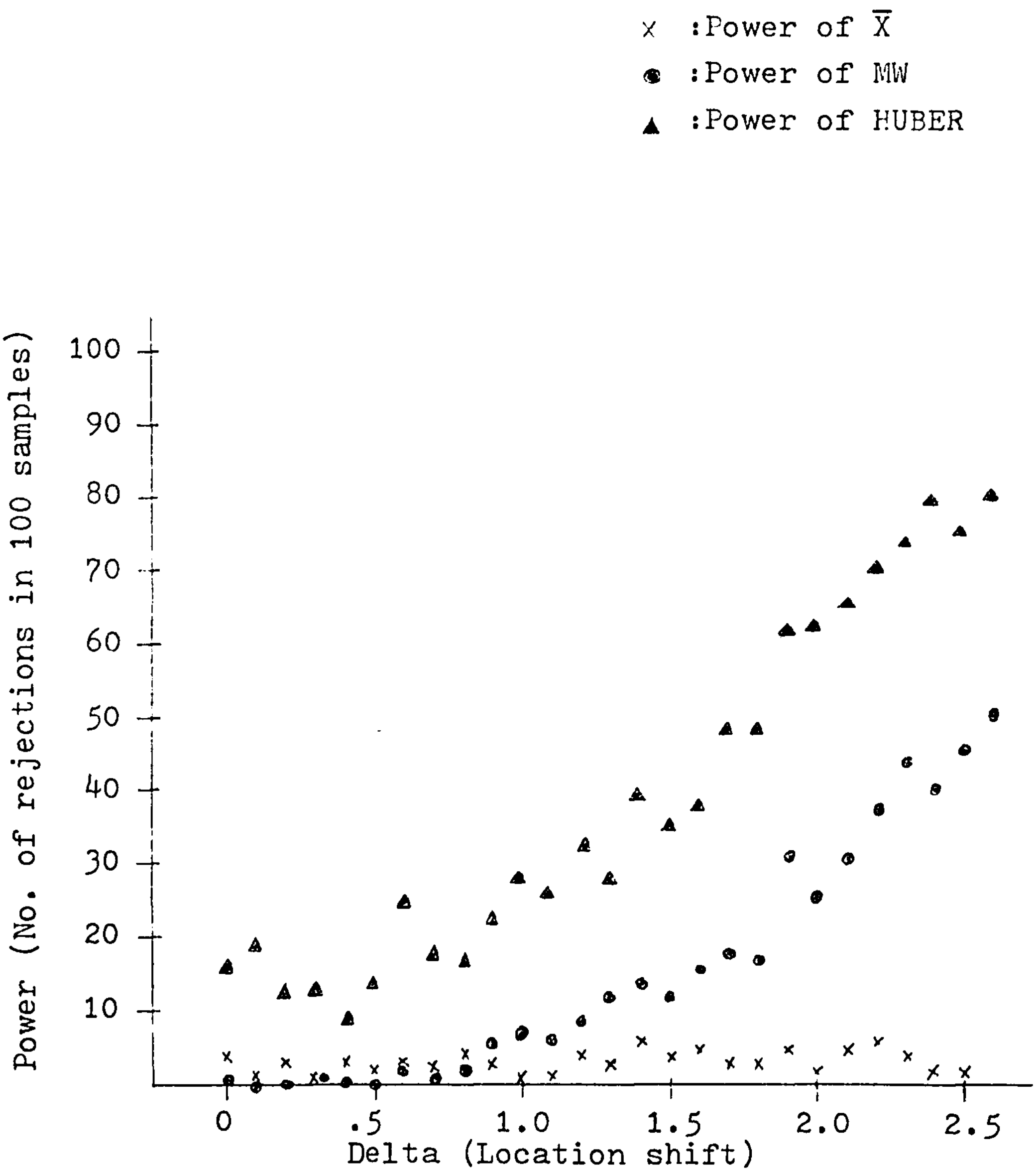

Fig.12, Compare of cauchy dist.
seed=7118863, n=5

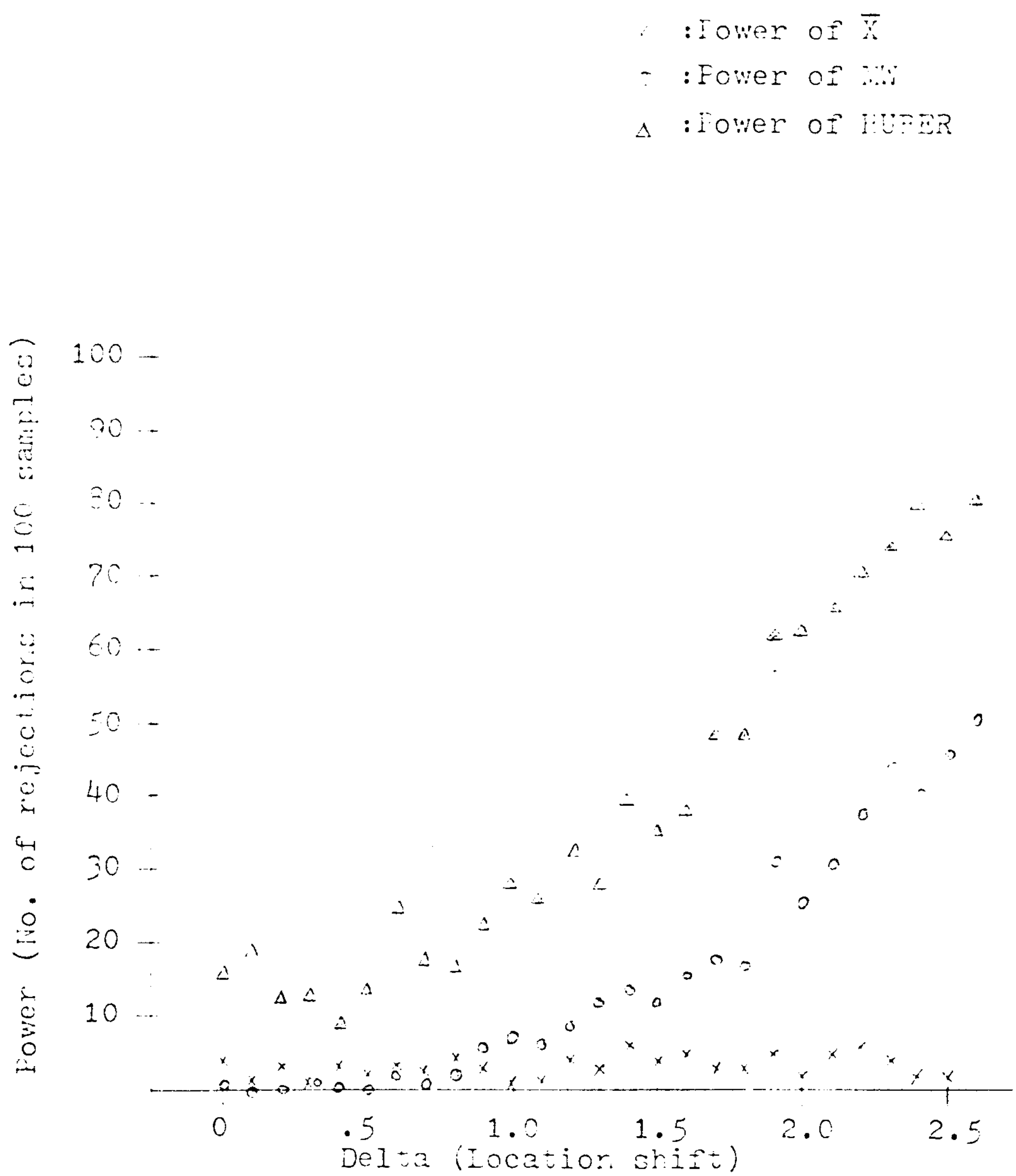

Fig.12, Compare of cauchy dist.
seed=7118863, n=5

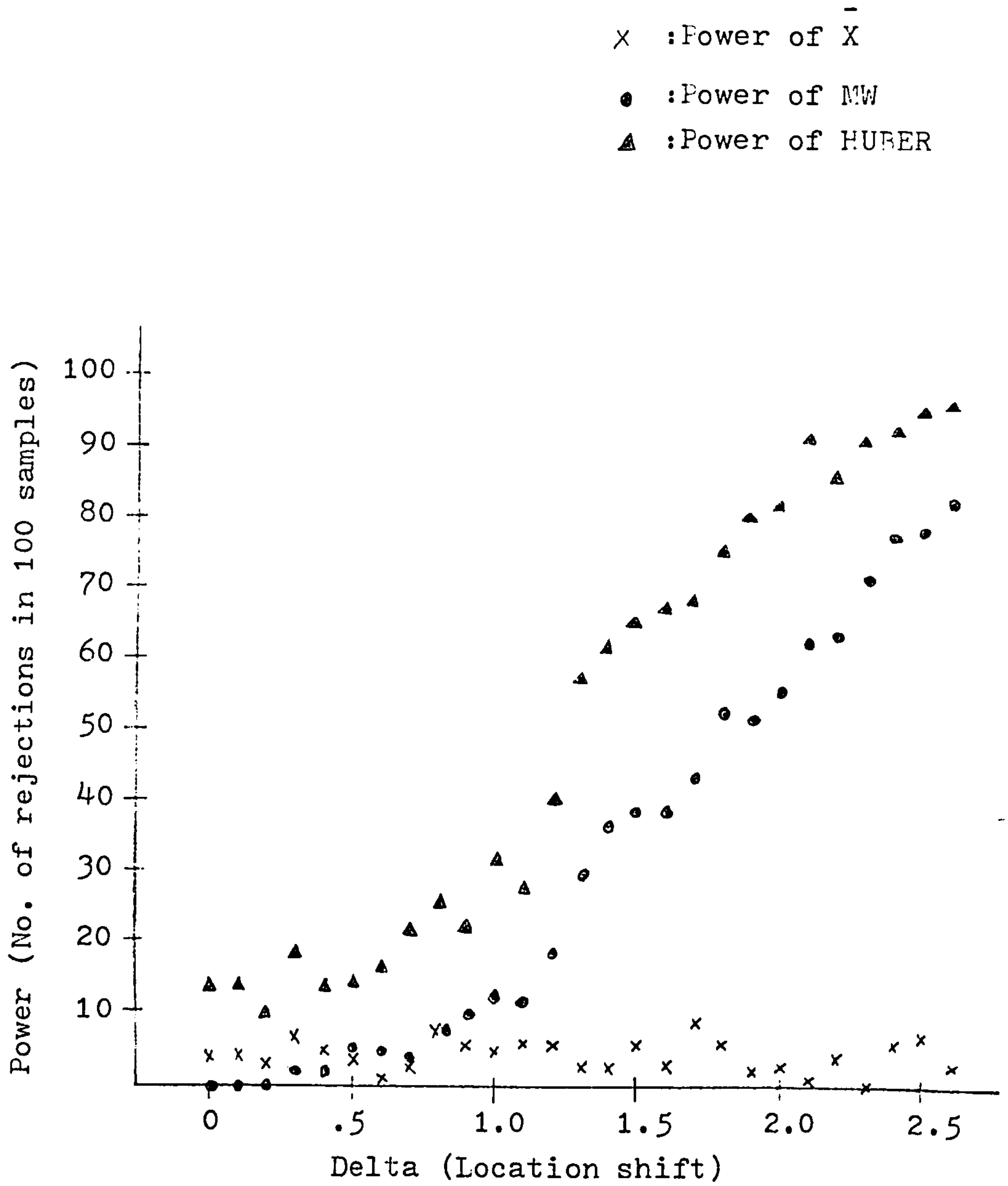

Fig.13, Compare of cauchy dist.
seed=7118863, n=10

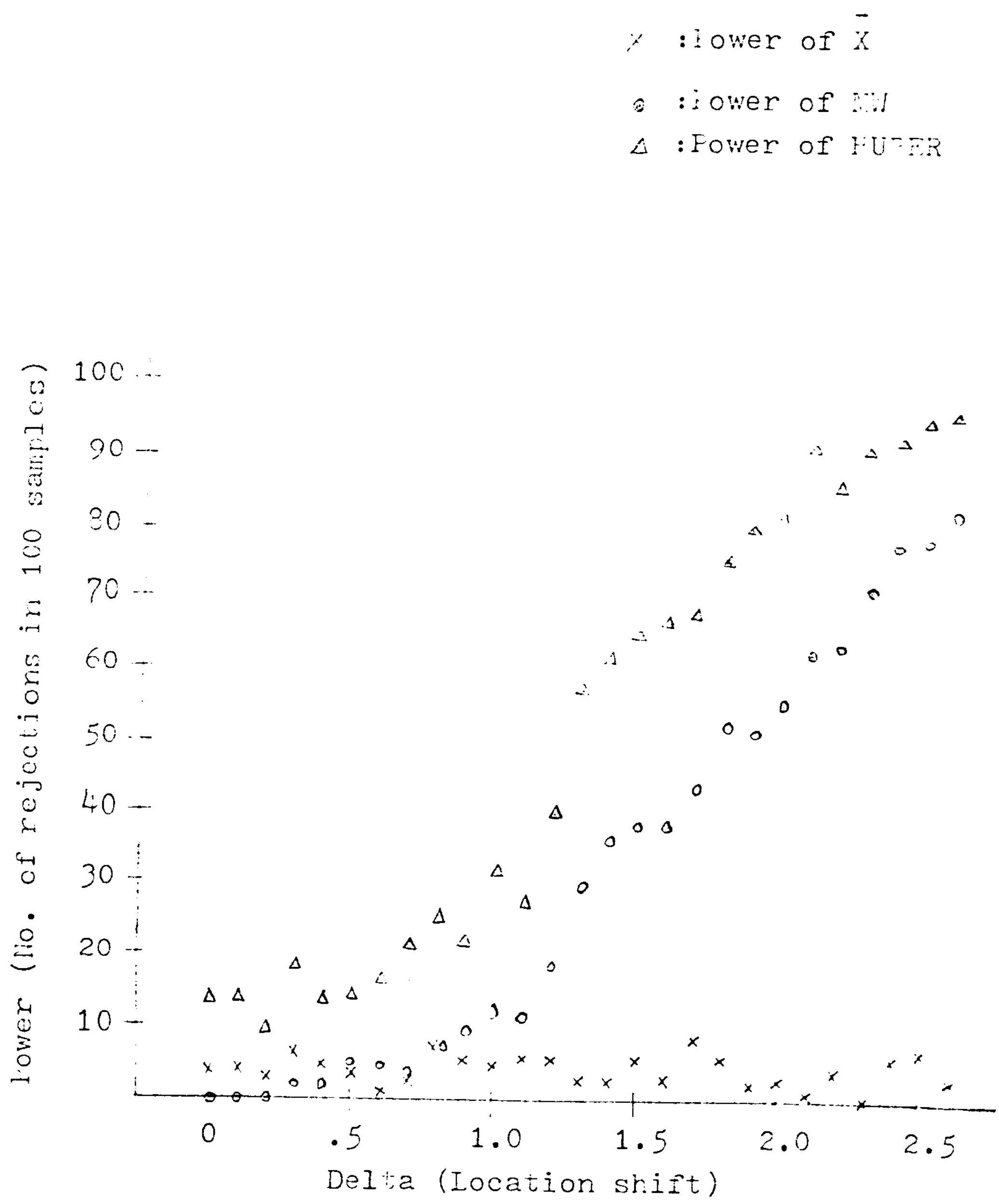

Fig.13, Compare of cauchy dist.
seed=7118863, n=10

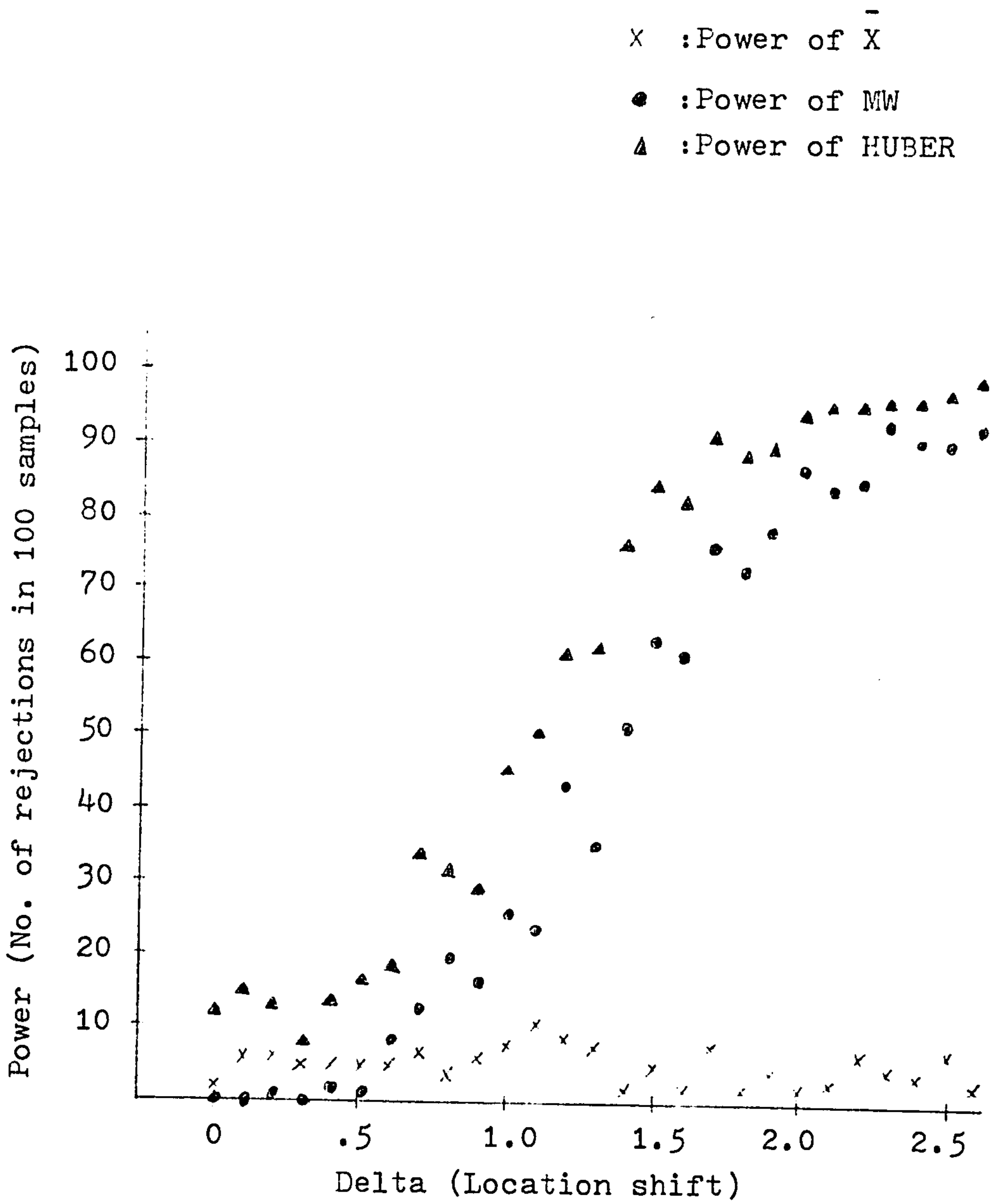

Fig.14, Compare of cauchy dist.
seed=7118863, n=15

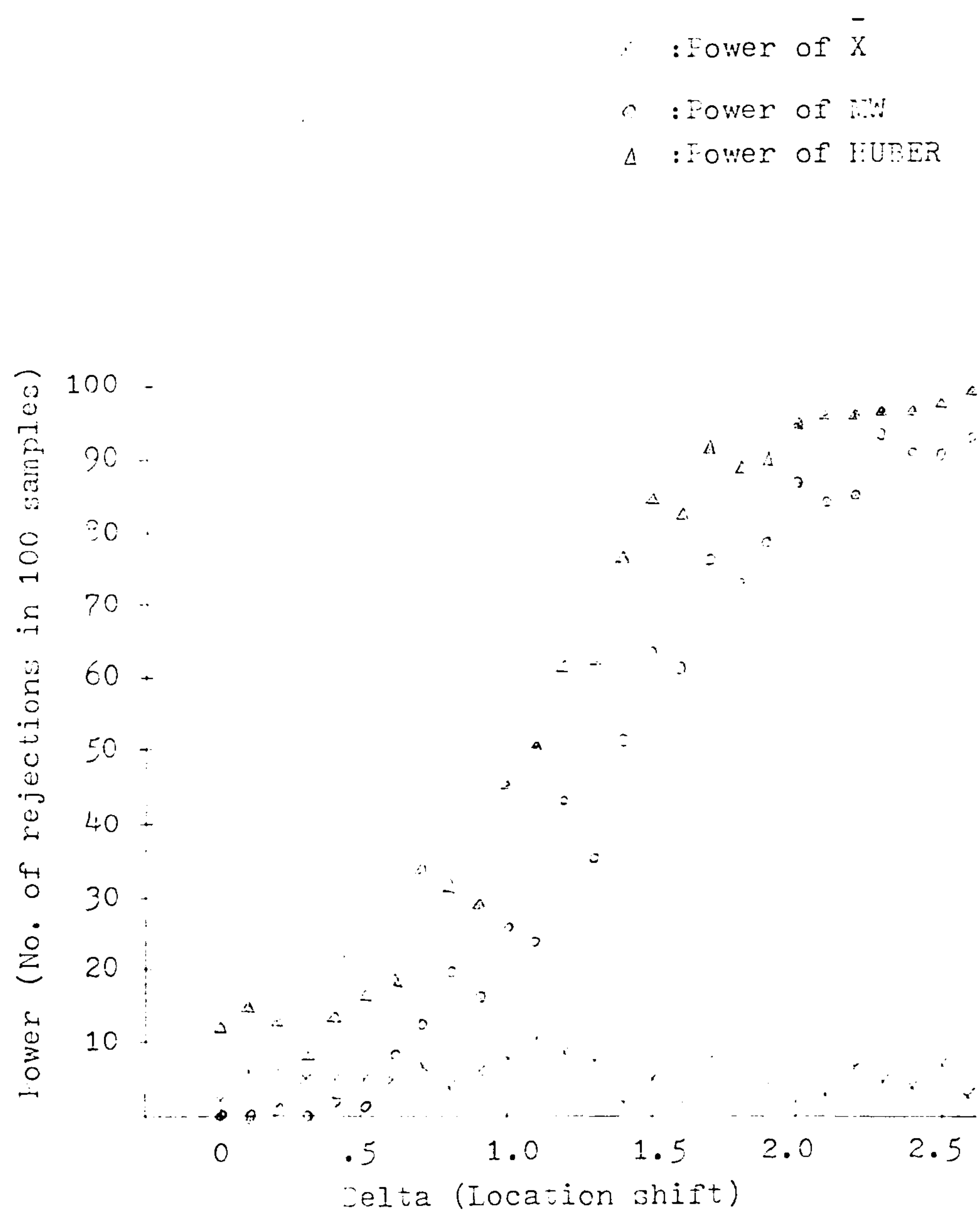

Fig.14, Compare of cauchy dist.
seed=7118863, n=15

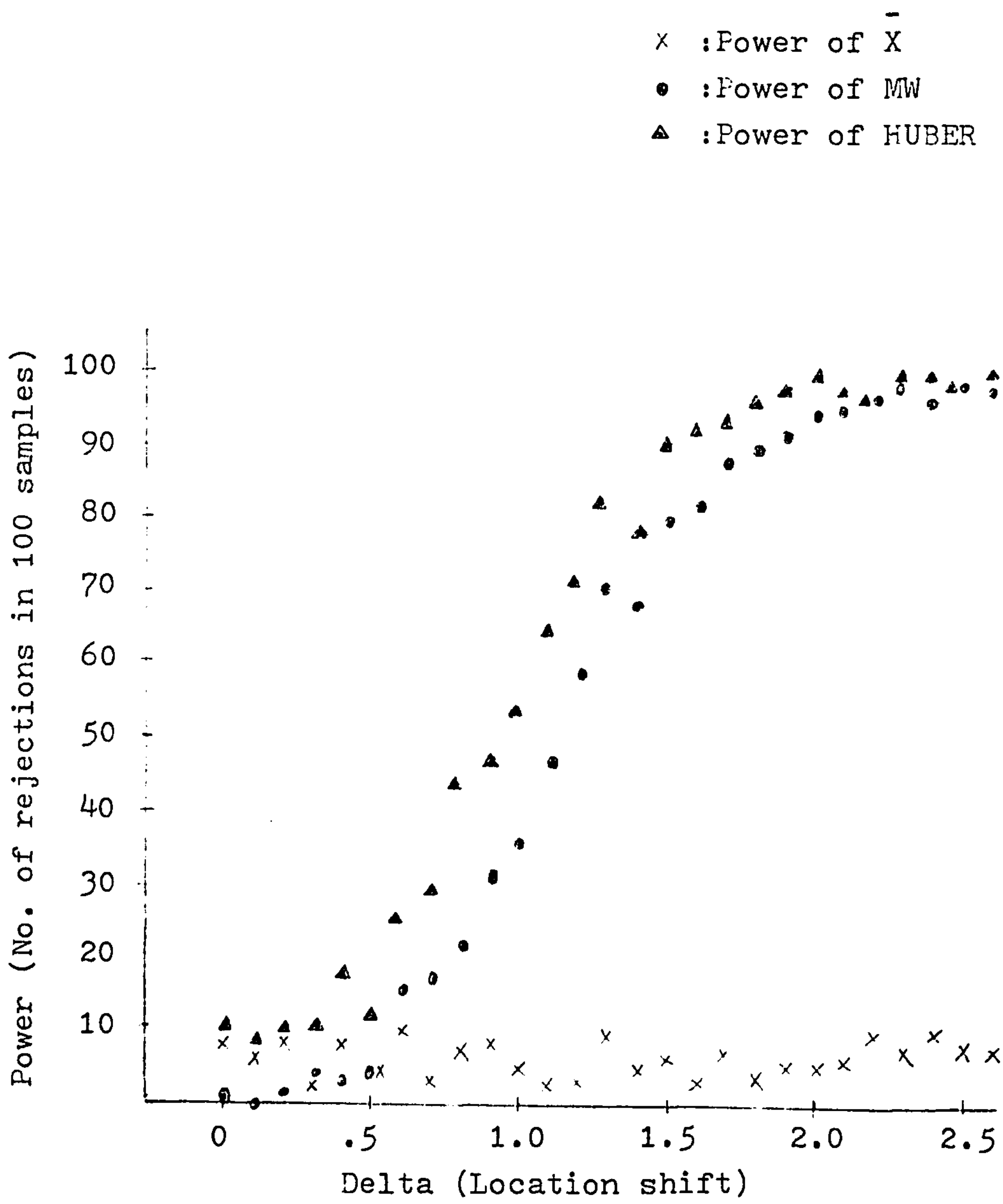

Fig.15, Compare of cauchy dist.
seed=7118863,n=20

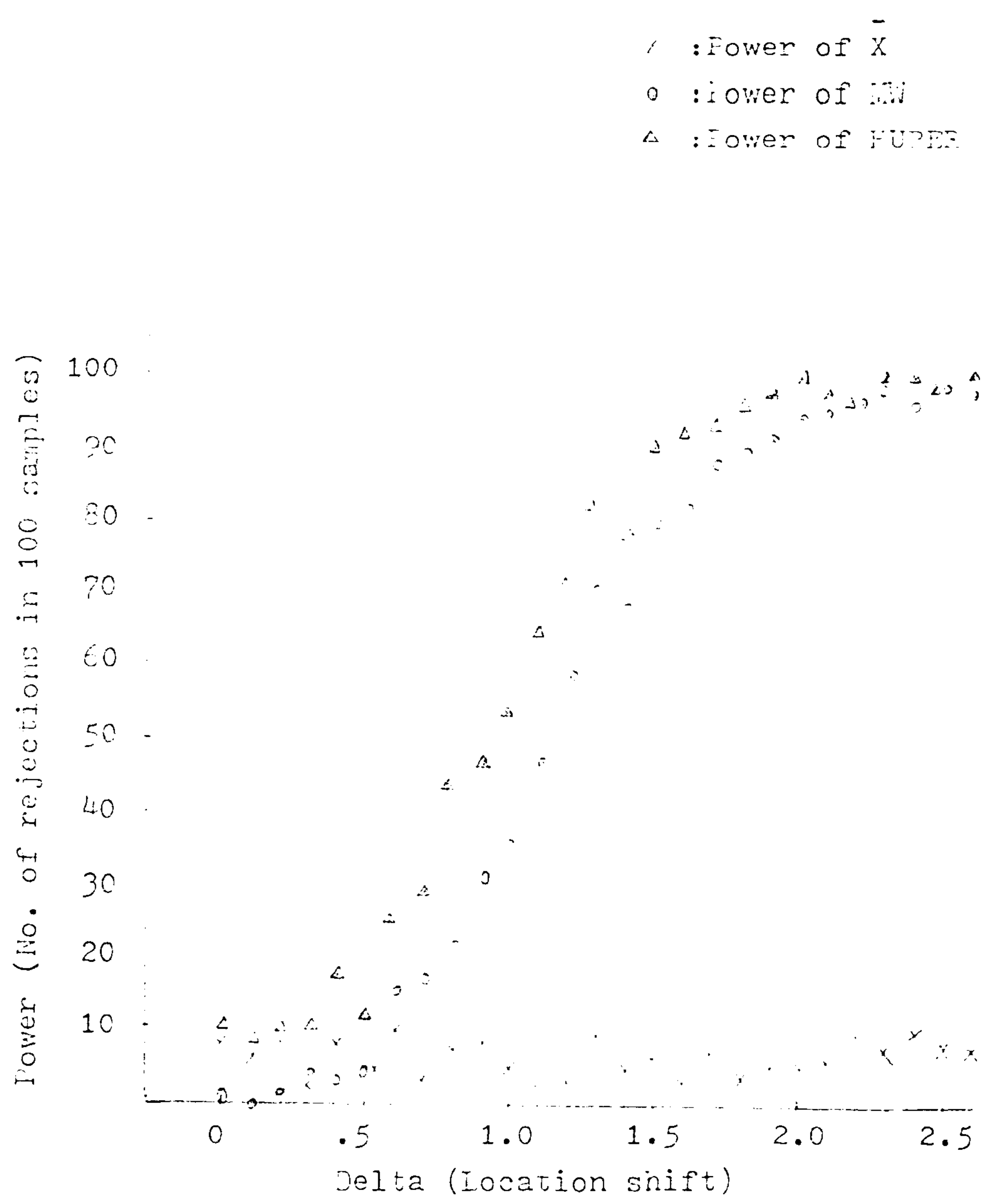

Fig.15, Compare of cauchy dist.
seed=7118863,n=20

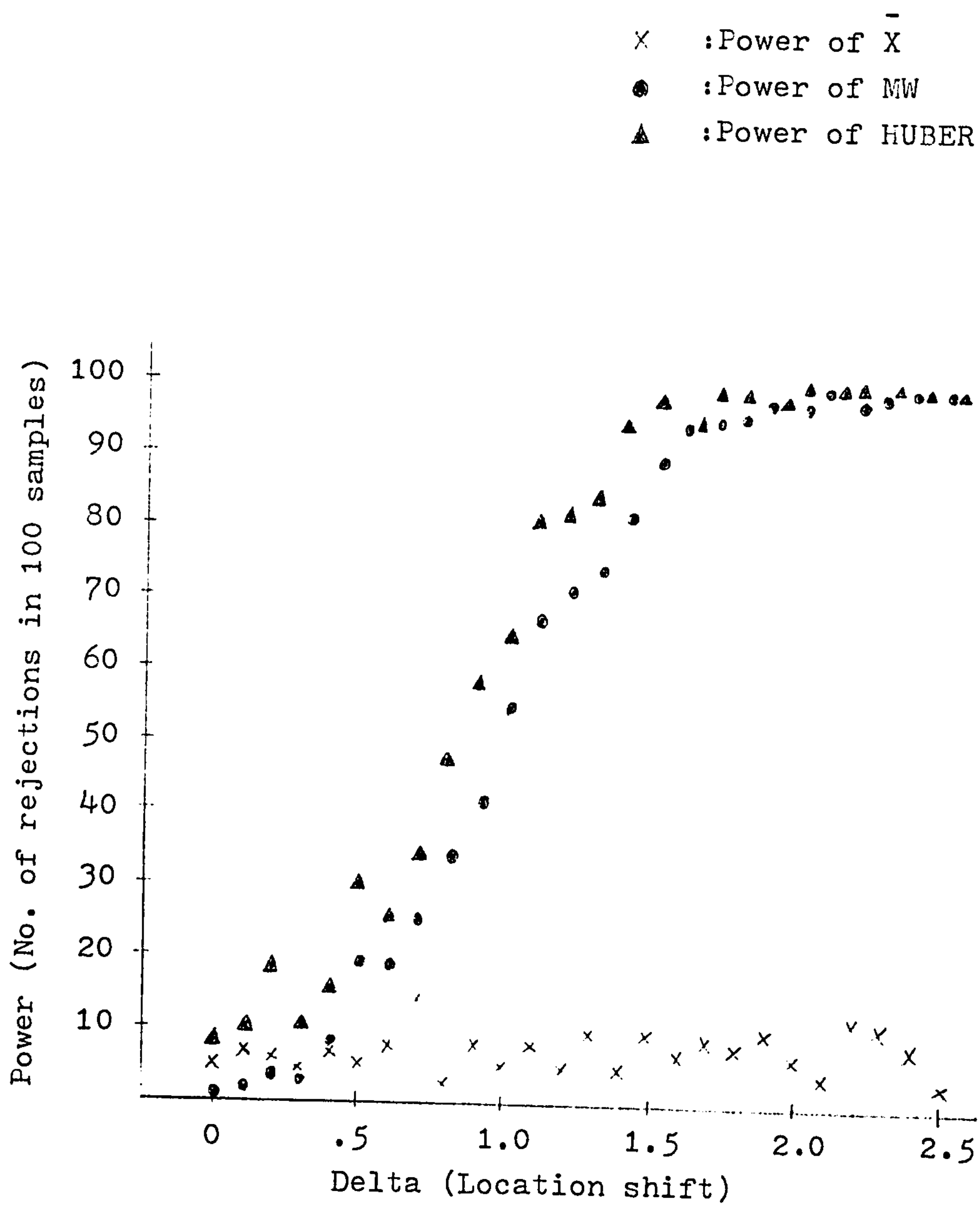

Fig.16, Compare of cauchy dist.
seed=7118863, n=25

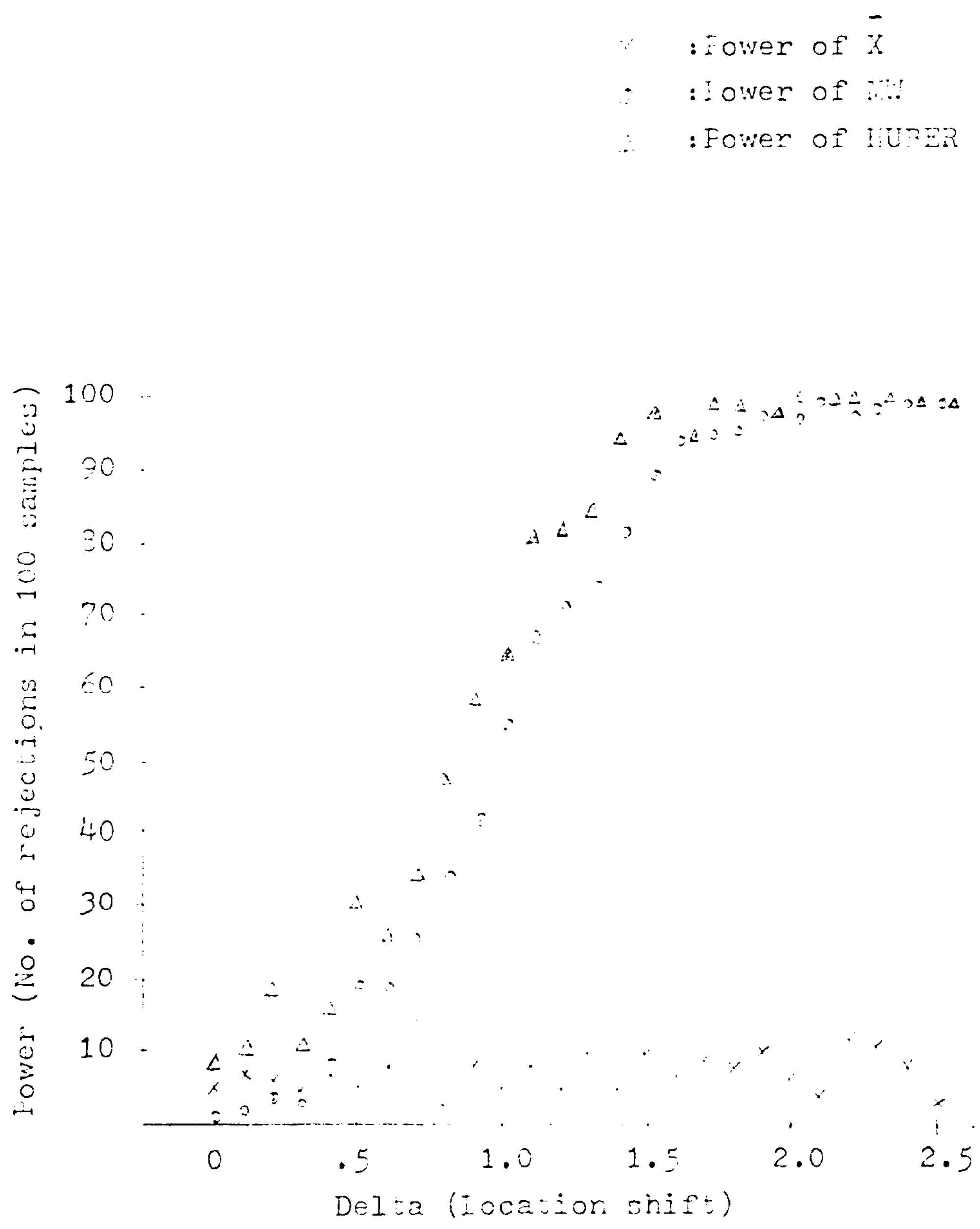

Fig.16, Compare of cauchy dist.
seed=7113863, n=25

Appendix I

TO GENERATE THREE ALTERNATIVE DISTRIBUTIONS

I.1

In order to generate contaminated normal distribution observations, the following assumptions are needed:

1. The mean of population distribution , μ=10.
2. The standard deviation σ=1. for normal condition
3. Contaminated scale H=3. or other assigned value
4. Probability of contaminated points p=.05 or other assigned value
5. Sample size n is assigned for monitoring the process

In I.M.S.L. subroutine library, we use subroutine GGNML to generate Gaussian deviation in interval (0,1) and GGUBS to generate uniform deviation in interval (0,1). Now , we can simulate contaminated normal observations using computer.

I.2

In order to generate cauchy distribution observations, the only assumption is that the median of distribution T=0. and , by using I.M.S.L. subroutine GGCAY to generate cauchy distribution observations.

I.3

In order to generate slash distribution observations, the following assumptions are needed:

1. The mean of population $\mu=10.$ for normal condition.

2. The standard deviation $\sigma=1.$ for normal condition.

Then, we can simulate slash distribution observations using computer.

The same way can be used to generate the sample observations with sample size n=any assigned value.

REFERENCE :

1. Acheson J. duncan, Quality control and industrial statistics. 4^{th} ed. 1974.

2. Hoaglin , Mostler, Tukey, Understanding robust and exploratory data analysis. 1983.

3. Mendenhall, Scheaffer, Wackerly, Mathematical statistics with applications. 2^{nd} ed. 1981.

4. Hamaker H. C. Acceptance sampling for percent deffective by variables and by attributies. Journal of quality technology. 1979.

5. Daugherty T. F. Two apparent contradictions in testing hypothesis. Journal of quality technology. 1980.

6. Rober L. Launer, Graham N. Wilkinson , Robustness in statistics. 1979.

7. Rober V. Hogg, An introduction to robust estimation. Dep. of statistics, U. of Iowa

8. Huber P. J. Robust estimation of location parameter. Annals of mathematical statistics, 35, 73-101. 1964.

9. Huber P. J. Robust statistics: a review . Annals of mathematical statistics, 43 , 1041-1067. 1972.

10. Huber P. J. Robust statistical procedures . Society for industrial and applied mathematics. 1977.

11. Huber P. J. Robust statistics. 1981.

12. Ronald H. Randles, Douglas A. Wolfe, Introduction to the theory of nonparametric statistics. 1979.

13. K. Sarkadi, I. Vincze, Methematical methods of statistical quality control. 1974.

14. Jean Dickson Gibbons, Nonparametric statistical inference. 1971.

15. D. f. Andrews, P. J. Bickel, F. R. Hampel, P.J. Huber, W. H. Rogers, J.W.Tukey, Robust estimates of location survey and advances. Princeton University Press,

Princeton, New Jersey, 1972.

16. A. S. C. Ehrenberg , Data Reduction, Analysing and interpreting statistical data. 1975.

VITA

PERSONAL	NAME : GON CHEN
	BIRTHDATE: SEPTEMBER 6 , 1950
	WIFE : W. C. HUANG
	CHILDREN : RU SHI AND RU HONG
EDUCATION	B. S. INDUSTRIAL ENGINEERING DEP. FONG CHIA UNIVERSITY , TAIWAN R. O. C.
PROFESSIONAL EXPERIENCE	CHUANG SHANG INSTITUTE OF SCIENCE AND TECHNOLOGY. INDUSTRIAL SAFTY ENGINEER 1/74 - 12/82